Analysis of Phage-type RNA Polymerase driven Transcription in *Physcomitrella patens* and *Arabidopsis thaliana*

DISSERTATION

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.) im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
der Humboldt-Universität zu Berlin

von

Diplom-Biologe Uwe Richter

Präsident der Humboldt-Universität zu Berlin
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I
Prof. Stefan Hecht, PhD

Gutachter: 1. Prof. Dr. Thomas Börner

2. Prof. Dr. Bruce Cahoon

3. Prof. Dr. Wolfgang Schuster

Tag der mündlichen Prüfung: 01. Februar 2011

Für Paul, Jovan und Katja

Summary

The relatively low coding capacity of plant mitochondrial and plastid genomes is in stark contrast to the highly complex transcription machinery found in these organelles. While eudicots such as *Arabidopsis* use the plastome encoded bacterial-type RNA polymerase together with at least two nuclear encoded RNA polymerases of the phage-type to transcribe the plastid genome, multiple phage-type RNA polymerases are also present in mitochondria. In the moss *Physcomitrella patens* three phage-type RNA polymerases PpRpoTm, PpRpoTmp1 and PpRpoTmp2 localise to mitochondria. The nuclear genes for PpRpoTmp1 and PpRpoTmp2 encode transit peptides with mitochondrial and plastid targeting properties suggesting dual localisation of these two enzymes. As shown in this study the third RpoT enzyme of *Physcomitrella* seems to be purely mitochondrial. The observation of multiple RNAPs in plant organelles raises the question of how these different transcriptional activities are facilitated and ultimately whether complexity of organellar transcription is a function of an increased demand for the regulation of mitochondrial and/or plastid gene expression.

This study therefore aimed to elucidate how the different transcriptional activities are facilitated in mitochondria of *Arabidopsis thaliana* and in both organelles of *Physcomitrella patens*. Insertional mutants for *PprpoTmp2* and *AtrpoTmp* were analysed in detail. As for *Arabidopsis RpoTm*, knock-out of *Physcomitrella RpoTmp1* was found to be lethal. Null mutant plants *PprpoTmp2* and *AtrpoTmp* show surprisingly similar but clearly convergent phenotypical aberrations reminiscent of phenotypes reported for other mitochondrial mutants.

In concordance with this observation steady-state transcript analysis in *AtrpoTmp* and *PprpoTmp2* revealed changes in the abundance of mitochondrial transcripts when compared to the wild type. While in *AtrpoTmp* decreased levels of specific transcripts were correlated with reduced abundances of the respiratory chain complexes I and IV, mitochondrial transcript abundance in *PprpoTmp2* was generally higher as in the wild type.

Even though lack of organellar transcriptional initiation in *PprpoTmp2* plants as compared to the wild type was not observed, eleven mitochondrial and nine plastid transcription initiation sites were mapped, providing the first verified data on plastid and mitochondrial transcription initiation in deep branching land plants.

Plastid promoter architecture in *Physcomitrella* recapitulates structures observed for plastid promoters identified in other land plants. Importantly all plastid TSS mapped in protonematal 'tissue' were found to be preceded by PEP like promoter motifs, suggesting a predominant role of PEP for plastid transcription in protonemata of *Physcomitrella*. Furthermore, mitochondrial core promoter motifs in *Physcomitrella* do not significantly deviate from the structure of mitochondrial promoters described in flowering plants, indicating that co-evolutionary diversification of the dedicated phage-type RNA polymerases with core promoter elements directly upstream of the TSS is of minor importance in the evolution of land plants. Moreover, evidence is provided that PpRpoTmp1 and PpRpoTmp2 are functional RNA polymerases, which both posses the inherent ability to recognize organellar promoters in a minimal *in vitro* transcription system without the aid of additional cofactors. The data suggest that coding for two RpoT proteins one representing an enzyme with a high portion of non-specific transcriptional activity, as seen for AtRpoTmp and PpRpoTmp1 and one that can act as a single-polypeptide enzyme and recognize numerous mitochondrial promoters *in vitro* as AtRpoTm and PpRpoTmp2 echo convergent inventions but reflect complementing roles of these RNA polymerases in plant mitochondrial transcription. The inability to isolate plants homozygous for T-DNA insertions in *AtRpoTm* provides direct evidence that the purely

mitochondrial RNAP is unconditionally required for the transcription of most mitochondrial genes in *Arabidopsis*. To the contrary, *Physcomitrella* RpoTmp2 knock out is not lethal and therefore indicates that its function can be partially substituted by one or both of the other RpoT proteins. Phenotypical aberrations of *rpoTmp2* plants however suggest that RpoTmp2 is important for normal growth and development. Whether the indicated mitochondrial dysfunction in *rpoTmp2* plants is directly due to the scarcity of a specific transcript as a consequence of lacking RpoTmp2 transcriptional activity, though no significant reduction was found among 32 mitochondrial transcripts tested, or due to an imbalance in the mitochondrial transcript pool remains open. Altered transcript levels in *AtrpoTmp* were found to result from gene-specific transcriptional changes, establishing that AtRpoTmp functions in distinct transcriptional processes within mitochondria. Decreased transcription of a specific set of mitochondrial genes in *AtrpoTmp* was not associated with changes in the utilisation of specific promoters. Therefore AtRpoTmp function is not promoter-specific but gene-specific. This indicates that additional gene-specific elements direct the transcription of a subset of mitochondrial genes by RpoTmp.

Zusammenfassung

Mitochondriale und plastidäre Genome kodieren eine relativ begrenzte Anzahl von Proteinen und ribosomalen RNAs. Gleichwohl besitzen pflanzliche Organellen hochkomplexe Transkriptionsapparate. Eudikotyle Pflanzen wie *Arabidopsis thaliana* etwa, nutzen zur Transkription des Plastidengenoms nicht nur die in der ptDNA kodierte bakterientypische RNA Polymerase, sondern greifen zusätzlich auf zwei im Kern kodierte phagentypische RNA Polymerasen zurück. Während in den *Arabidopsis* zwei mitochondriale kernkodierte RNA Polymerasen vom Phagentyp nachgewiesen werden konnten, existieren in Mitochondrien des Bryophyten *Physcomitrella patens* drei phagentypische RNA Polymerasen. Die nukleären Gene *PpRpoTmp1* und *PpRpoTmp2* kodieren N-terminale Signalpeptide, die eine duale Lokalisation in Mitochondrien und Plastiden von *Physcomitrella* vermitteln. Wie in der hier vorgelegten Arbeit gezeigt, kodiert *PpRpoTm* hingegen für ein rein mitochondriales Protein. Das Auftreten multipler Transcriptionsaktivitäten in den Organellen von Landpflanzen wirft die Frage auf, wie die verschiedenen Polymerasen zur Transkription von mt- und ptDNA beitragen. In der vorliegenden Arbeit wurde daher der spezifische Einfluss der verschiedenen kernkodierten phagentypischen RNA Polymerasen auf die organelläre Genexpression in *Physcomitrella* und die mitochondrial Genexpression in *Arabidopsis* untersucht. Während das Fehlen von AtRpoTm in *Arabidopsis* und PpRpoTmp1 in *Physcomitrella* lethal ist, konnten Insertionsmutanten für *PpRpoTmp2* und *AtRpoTmp* einer detailierten Untersuchung unterzogen werden. Sowohl *PprpoTmp2*, als auch *AtrpoTmp* Pflanzen zeigten Abweichungen im Phänotyp charakteristisch für eine mitochondriale Dysfunktion. In Übereinstimmung mit einer Störung der mitochondrialen Funktion, konnten in beiden Mutanten Abweichungen in der Abundanz mitochondrialer Transkripte im Vergleich zum Wildtyp nachgewiesen werden. Während in *AtrpoTmp* verminderte Quantitäten spezifischer Transkripte mit der Reduktion der Abundanz der respiratorischen Komplexe I und IV korrelierte, waren die Transkriptmengen für die überwiegende Zahl mitochondrialer Gene in *PprpoTmp2* Pflanzen erhöht. Der Ausfall der Transkriptionsaktivitäten PpRpoTmp2 und AtRpoTmp führte weder in *P.patens* noch in *A.thaliana* zu einem vollständigen Ausfall der Transkriptionsinitiation an organellären Promotoren.

Die Identifizierung von 11 mitochondrialen und 9 plastidären Transkriptionsinitiationsstellen in *Physcomitrella patens* lieferte jedoch die ersten verifizierten Daten bezüglich der Promotorstruktur beider organellärer Genome in einer früh abzweigenden Linie der Landpflanzen. Des weiteren konnten die identifizierten Promotoren zum Nachweis der Transkriptionsaktivität von PpRpoTmp1 und PpRpoTmp2 in einem *in vitro* Transkriptionssystem herangezogen werden. Für beide Proteine konnte eine inhärente Fähigkeit zur Promotorerkennung ohne zusätzliche Kofaktoren nachgewiesen werden. Während PpRpoTmp2 an verschiedenen Promotoren die Transkription initiierte, zeigte RpoTmp1, trotz hoher basaler Transkriptionsaktivität, eine weit niedrigere Promotorspezifität.

Die hier vorgestellten Studien unterstreichen die essentielle Bedeutung von AtRpoTm und PpRpoTmp1 für die Transkription mitochondrialer bzw organellärer Gene in *Arabidopsis* und *Physcomitrella*. Im Gegensatz dazu können die Funktionen von AtRpoTmp und PpRpoTmp2 partiell durch andere organelläre RNAPs ersetzt werden. Phänotypische Abweichungen belegen jedoch, das AtRpoTmp und PpRpoTmp2 für die normale Entwicklung von *Arabidosis* bzw. *Physcomitrella* Pflanzen essentiell sind. Während eine Reihe von mtDNA kodierten Transkripten überakkumulierten konnte eine Reduktion von Transkriptmengen in *PprpoTmp2* Pflanzen für keines der getesteten mitochondrialen Transkripte nachgewiesen

werden. Ein gestörtes Gleichgewicht auf der Ebene der Transkriptakkumulation in *PprpoTmp2* Mitochondrien könnte daher als ursächlich für die beobachteten Abweichungen im Phänotyp der *PprpoTmp2* Pflanzen interpretiert werden.

Veränderte mitochondriale Transkriptmengen in *AtrpoTmp* Pflanzen korrelierten hingegen deutlich mit genspezifischen Änderungen auf der Ebene der mitochondrialen Transkription. AtRpoTmp muss daher als essentiel für die normale Expression eines spezifischen Sets mitochondrialer Gene angesehen werden. Jedoch konnten für dieses spezifische Set mitochondrialer Gene keine AtRpoTmp spezifischen Promotormotive mit reduzierter Aktivität identifiziert werden. Statt dessen waren die Initiationsraten an allen Promotoren stromaufwärts von mitochondrialen Genen mit reduzierten Transkriptmengen reduziert.

Die Funktion von AtRpoTmp ist daher nicht promotorspezifisch, sondern genspezifisch vermittelt. Es erscheint daher wahrscheinlich, daß für einen Teil der mitochondrialen Gene genspezifische Elemente existieren, welche die Transkription durch AtRpoTmp dirigieren.

Table of contents

ABBREVIATION

A	adenine
aa	amino acid
Abbr.	abbreviation
(d)ATP	(deoxy) adenosine triphosphate
BLAST	Basic alignment search tool
BSA	bovine serum albumin
bp	base pair
cDNA	complementary DNA
cRNA	complementary RNA
C	cytosine
°C	degrees celsius
Ci	curie
cp	chloroplast
CTAB	cetyl trimethyl ammonium bromide
(d)CTP	(deoxy) cytidine triphosphate
d	day(s)
DNA	deoxyribonucelic acid
DMF	N,N-Dimethylformamide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
Fig.	figure
G	guanin
g	gravitational acceleration
G	guanine
(d)GTP	(deoxy)guanosine triphosphate
h	hour(s)
His	histidin
kb	kilobasepairs
kDa	kilodalton
l	liter
MES	morpholinoethan-sulfonic acid
min	minutes
MOPS	morpholinopropan-sulfonic acid
mRNA	messenger RNA
mt	mitochondrial
MWCO	molecular weight cut-off
NBT	nitro blue tetrazolium
NEP	nucleus encoded plastid RNAP
nt	nucleotide
(d)NTP	(deoxy) nucleoside triphosphate
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEP	plastid encoded plastid RNAP

pH	$-\log_{10}[H_3O^+]$
PMSF	phenylmethyl sulphonyl fluoride
RACE	rapid amplification of cDNA ends
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	rotations per minute
RpoT	phage T7/T3-like RNAP
rRNA	ribosomal RNA
RT	reverse Transkriptase
SDS	sodium dodecyl sulfate
T	thymin
Tab.	table
TAP	tobacco acid pyrophosphatase
Tris	tris (hydroxymethyl)-aminomethane
tRNA	transfer RNA
dTTP	deoxythymidin triphosphate
Trx	thioredoxin
U	unit(s)
UTP	uridin triphosphate
UV	ultra violet
v/v	volume per total volume
w/v	weight per volume
WT	Wild type

1 INTRODUCTION

1.1 Chloroplasts and mitochondria are endosymbiont-derived organelles

Mitochondria and plastids are descendants of bacterial endosymbionts (Margulis 1970, Margulis 1981) and both possess their own reduced genomes. Analyses of the mitochondrial DNA (mtDNA) infer single ancestry for mitochondria with the closest contemporary relatives found within the α-proteobacterial order of Rickettsiales (Williams *et al.* 2007). Recent studies of unicellular eukaryotes have raised the possibility that the origin of mitochondria might be tightly bound to the origin of Eukaryotes, since eukaryotic cells might have originated in an endosymbiotic event of an archebacterial-like and an α-proteobacterial ancestor (Martin and Müller 1998). It is to mention that this hypothesis is still fiercely debated among cell evolutionists. Unlike mitochondria, which appear to have originated in the common ancestor of all known extant eukaryotic lineages, the plastid arose sometime later within a well-defined subgroup of eukaryotes. The origin of chloroplasts traces back to a primary endosymbiotic event between an eukaryotic host and a relative of extant cyanobacteria representing the root of the plant kingdom (Rodriguez-Ezpeleta *et al.* 2005), which comprises three major algal groups, glaucophytes, red algae and green algae, and their plant descendants all containing primary plastids, delimited by two membranes.

During the roughly two billion years since eukaryotes arose, the majority of the genes from the ancestral mitochondrial and plastid genomes were either relocated to the nuclear genome or lost relatively early in the process of both endosymbiotic events (Gray *et al.* 1999), but transfer of genetic material from the organellar genomes to the nucleus is clearly a continuing process since recently transferred organelle DNA fragments are found in a wide variety of eukaryotes (Brennicke *et al.* 1993, Wischmann and Schuster 1995). The nuclear genome of *Arabidopsis thaliana*, for instance, contains a large insert of 620 kb of mtDNA on chromosome 2 (Lin *et al.* 1999, Stupar *et al.* 2001), as well as 17 fragments of cpDNA totalling 11 kb (The Arabidopsis Genome Initiative 2000), with sequence similarity scores indicative of rather recent transfer events. Furthermore, functional gene transfer from the tobacco plastid genome to the nucleus was shown to be inducible reproducing a key process in the evolution of eukaryotic cells in laboratory experiments (Stegemann and Bock 2006). The majority of components necessary for proper function of the diverse mitochondrial and plastid metabolic pathways and their underlying genetic processes are encoded in the nucleus. After synthesis in the cytosol proteins must be imported into organelles (Herrmann 2003, Peeters and Small 2001). As a direct consequence unidirectional functional gene transfer is dependent on the acquisition of functional transit peptides redirecting the gene products into the proper compartment, where subsequently the function of the original still encoded in the organelle might be usurped. Since this is a stochastic process, i.e., without directed re-routing of the gene products back to the donor organelle, proteins were also 'mistargeted'. Therefore, this process allowed for the acquisition of new functions in different compartments of the eukaryotic cell. It is chance, natural selection and lineage diversification including drift that determine the targeting fate of genes that organelles donate to the genomes of their hosts. This led to observable whole pathway transfers between compartments (Martin and Schnarrenberger 1997). For biochemical pathways redundantly encoded in both the host and its endosymbiont, competition can arise (Martin and Herrmann 1998, Martin and Schnarrenberger 1997). In such a scenario, the pathway of the symbiont might dominate (Lange *et al.* 2000, Tovar *et al.* 2003), but also hybrid pathways can be the result (Aravind *et al.* 2003, Martin and Schnarrenberger 1997, Schnarrenberger and Martin 2002). Another

important consequence is that some protein complexes or pathways are hybrids of nuclear and organelle encoded proteins as seen for the respiratory chain complexes of mitochondria (Berry 2003). Altogether endosymbiosis resulted in functional redundancy eliminated through differential gene loss. Gene products functionally or structurally 'homologous' repeatedly replaced pre-existing nuclear encoded counterparts (Sanchez *et al.* 1996). Progressive reduction and integration with the new host resulted in the highly specialised organelles existing today (Kutschera and Niklas 2005).

1.2 Transcription in endosymbiotically derived cell organelles

1.2.1 Mitochondrial genomes

All eukaryotes that have been studied thus far bear derivatives of the α-proteobacterial endosymbiont, and for most the mitochondrion has a central role in energy metabolism as well as in providing several other catabolic and anabolic functions (Gray *et al.* 2004). Mitochondria containing their own genomes (chondromes) are found in all but a few anaerobic eukaryotes (Gray *et al.* 1999). The jakobid flagellate *Reclinomonas americana* harbors the least derived mitochondrial genome characterised to date, containing at least 98

Table 1: Sizes and coding content of some organelle and prokaryote genomes

Genome	Length (kbp)	Number of genes	Encoded proteins	Refseq.
α-proteobacteria				
Caulobacter crescentus	4017		3767	NC_002696
Mesorhizobium loti	7596		7281	NC_002678
Bradyrhizobium japonicum	~9100		~8300	NC_004463
other eukaryotes				
mt *Reclinomonas americana*	69	98	67	NC_001823
mt *Saccharomyces cerevisiae*	86	43	19	NC_001224
mt *Homo sapiens*	17	37	13	AC_000021
mt *Cyanidioschyzon merolae*	32	62	34	NC_000887
Chlorophyceae				
mt *Chlamydomonas reinhardtii*	16	25	8	NC_001638
mt *Scenedesmus obliquus*	43	53	20	NC_002254
Streptophyta				
mt *Chaetosphaeridium globosum*	57	77	46	NC_004118
mt *Chara vulgaris*	68	76	46	NC_005255
mt *Klebsormidium flaccidum*	100	nd.	nd.	OGMP
mt *Marchantia polymorpha*	187	74	41	NC_001660
mt *Physcomitrella patens*	105	69	42	NC_007945
mt *Cycas taitungensis*	415	68	39	NC_010303
mt *Arabidopsis thaliana*	367	56	33	NC_001284
mt of certain *Cucurbitaceae*	2400	nd.	nd.	Ward, et al., 1981
Cyanobacteria				
Synechocystis sp.	3573		3168	NC_000911
Prochlorococcus marinus	1660		1884	NC_005071
Nostoc PCC 7120	6413		5368	NC_003272
other Algae				
cp *Cyanidioschyzon merolae*	150	243	207	NC_004799
cp *Scenedesmus obliquus*	161	113	77	NC_008101
cp *Chlorella vulgaris*	151	210	174	NC_001865
cp *Chlamydomonas reinhardtii*	204	109	69	NC_005353
Streptophyta				
cp *Chaetosphaeridium globosum*	131	141	98	NC_004115
cp *Chara vulgaris*	185	148	105	NC_008097
cp *Marchantia polymorpha*	121	134	89	NC_001319
cp *Physcomitrella patens*	123	130	85	NC_005087
cp *Pinus thunbergii*	120	117	69	NC_001631
cp *Nicotiana tabacum*	156	144	98	NC_001879
cp *Arabidopsis thaliana*	154	129	85	NC_000932
cp *Zea mays*	140	122	76	NC_001666
Non-phosynthetic plastids				
cp *Epifagus virginiana*	70	71	25	NC_001568
cp *Toxoplasma gondii RH*	35	63	26	NC_001799
cp *Eimeria tenella*	35	65	28	NC_004823

Organelle **G**enome **M**egasequencing **P**rogram (OGMP) at http://www.bch.umontreal.ca/ogmp/

genes encoded on a circular chromosome (Lang *et al.* 1997). The typical mitochondrial genome encodes 40–70 genes arranged on either circular or linear chromosomes of 15–60 kb (see Tab. 1). In Trypanosomes, diplomonades and apicomplexans, the chondrom has been modified substantially, resulting in extremes in genome structure reviewed by Jackson and colleagues (Jackson *et al.* 2007).

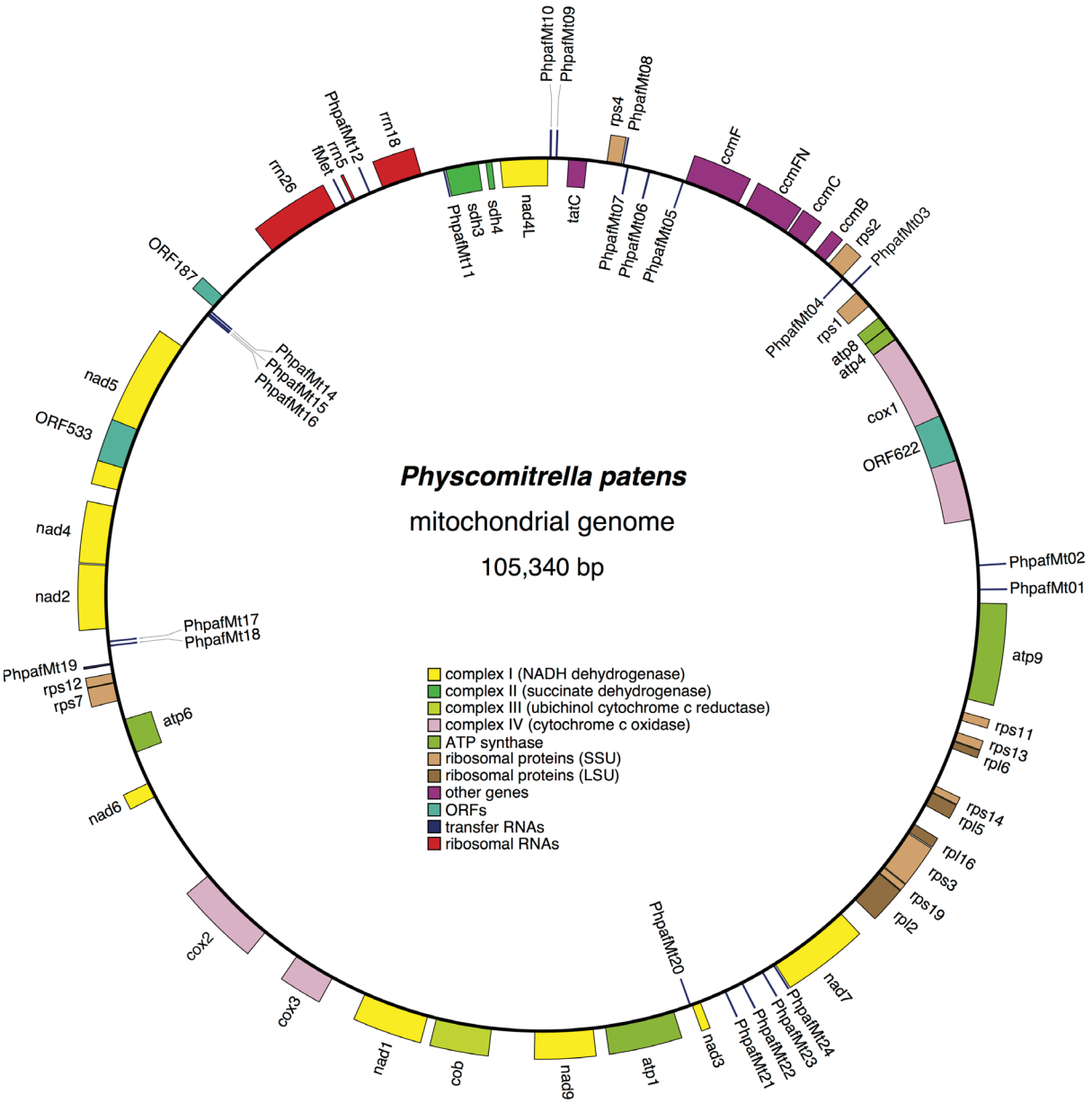


Figure 1: Circular map of the mitochondrial genome of *Physcomitrella patens*. Genes encoded by the counterclockwise strand are shown and indicated outside the circle, whereas genes encoded by the clockwise strand are inside. The map was created by OGDRAW (Lohse *et al.* 2007)and embellished in Adobe Illustrator.

Plant mitochondrial genomes expanded secondarily in length and range from 60 to thousands of kb. Although considerably varying in size they contain a relatively stable number of 56 to 77 genes (Timmis J.N. 2004), exemplified by the mtDNA of *Arabidopsis thaliana* (~367 kbp) being 20 times larger than the human mitochondrial genome (~17 kbp) but encoding only

approximately twice the number of proteins (Anderson *et al.* 1981, Unseld *et al.* 1997). Thirty-three proteins, three rRNAs and 20 tRNAs are encoded by the *Arabidopsis* mtDNA (Duchene and Marechal-Drouard 2001, Unseld *et al.* 1997), whereas the *Arabidopsis* mitochondrion contains up to 2000 proteins (H. Millar, personal communication). Recently, Terasawa and collegues reported that the mitochondrial genome of the leafy moss *Physcomitrella patens* (see Fig.1) representing a member of deep branching land plants does not show the multipartite structure of the mitochondrial genomes found in flowering plants like *Arabidopsis*. It contains genes for three rRNAs (*rrn23*, *rrn18*, and *rrn5*), 24 tRNAs, and 42 conserved mitochondrial proteins. *Physcomitrella* retains prototype features among land plant mitochondrial genomes and expansion in size is much less pronounced when compared with angiosperms and the mitochondrial genome structure appears more similar to *Chara vulgaris*, *Chaetosphaeridium globosum*, and *Marchantia polymorpha* (Terasawa *et al.* 2007). Significant synteny with angiosperm and chlorophyte mitochondrial genomes was not detected.

In general, streptophyte mitochondrial genes are dispersed or organised in clusters, mainly encoding the components of the respiratory chain and of the translational apparatus, a structure also found in yeast mitochondrial genomes (Burger *et al.* 2003). The mtDNA of *Chlamydomonas reinhardtii* is thought to be evolving more rapidly than the nuclear DNA, similar to the situation described in mammals (Brown *et al.* 1982, Pesole *et al.* 1999). This is in stark contrast to the situation in land plants (Muse 2000, Wolfe *et al.* 1987). Large intergenic regions have been extended through duplications as well as by incorporation of introns and 'foreign' DNA segments from the other compartments and of unknown origin (Marienfeld *et al.* 1999, Palmer *et al.* 2000). The evolution of mtDNA sequence is slow due to low mutation rates (Palmer 1990). The fluidity of the intergenic regions most pronounced in mitochondria of spermatophytes, together with low mutation rates of land plants might be a driving factor for the gain of complexity seen in organelle RNA metabolism, specifically for the RNA transcription and maturation machineries (Maier *et al.* 2008). Yet another consequence of the high recombinational activity is the appearance of ORFs, which contain pieces of known genes and unassigned sequences. Expression of those chimeras is found to be associated with cytoplasmic male sterility (CMS, Schnable and Wise 1998).

In addition to the relatively large and complex plant mitochondrial genome a few plant and some fungal mitochondria contain small extrachromosomal DNAs apparently absent from animal cells. They are found in two forms circular and linear and are described as replicons or plasmids due to the observation that they seem to replicate autonomously. The circular form is small in size, usually 1 to 2 kb, not coding for ORFs of significant size and it is found in a very diverse range of plant species (Lonsdale and Grienenberger 1992). The latter peculiarity is shared with the other form, the linear plasmids. Linear extrachromosomal DNAs in plant mitochondria, structurally very similar to their fungal counterparts, are larger, ranging from 2 to 13 kb and bear several expressed ORFs. Functional relevance of these ORFs for plant development is indicated by experiments showing differential protein expression in *Brassica napus* (Handa *et al.* 2002). Both plant and fungal linear plasmids usually contain terminal inverted repeats and encode DNA and/or RNA polymerases possibly to drive expression and replication in an autonomous manner. The invertron structure also found in some DNA viruses like phi29 and adenoviruses suggest a viral origin of plant and fungal linear plasmids. Furthermore, structural variations between plant linear plasmids and their unbalanced distribution strongly indicate diverse origins rather than common ancestry inside of the plant kingdom. To date the most likely scenario is that linear plasmids found in only nine plant species so far were acquired multiple times via horizontal gene transfer, possibly from fungal

phyto-pathogens (Handa 2008, McDermott *et al.* 2008). In general, plant organelle genes contain introns belonging to the group I and group II families (reviewed in Bonen and Vogel 2001). Interestingly, the most parsimonious explanation for the origin of the group I intron found in mitochondrial *cox1* of 48 angiosperm genera requires 32 independent horizontal transfer events also from a fungal source (Cho and Palmer 1999), underlining the importance of lateral and horizontal gene transfer for the evolution of plant mitochondrial genomes.

1.2.2 Plastid genomes

The origin of plastids is rooted within the endosymbiotic uptake of a photoautotrophic cyanobacterium some 1.5 billion years ago (see 1.1). They facilitate a variety of essential metabolic roles such as synthesis of fatty acids, tetrapyrroles, and aromatic compounds besides their central role in the fundamental biochemical process of photosynthesis. In 'higher plants' different types of plastids specialised to execute specific functions in different tissues exist, besides chromoplasts, amyloplasts, etioplasts and gerontoplast the most prominent and thoroughly studied being the photosynthetically active chloroplast of green tissues. Plastids of plants belonging to the taxon *Streptophytina*, including *Charales*, *Coleochaetales* and land plants have genomes that show significantly less structural diversity than mitochondrial genomes. Most of their plastid genomes contain a very stable number of 100 to 120 genes (see Tab. 1), encoded on mono- and multimeric circles (Bedbrook and Bogorad 1976, Herrmann *et al.* 1975, Kowallik and Herrmann 1972, Lilly *et al.* 2001), as well as on linear molecules (Oldenburg and Bendich 2004). Even though the ratio between those is still a matter of discussion (Bendich 2004), the majority of data underline the importance of the circular composition as a subset of plastid genomes in a variety of land plants (Backert *et al.* 1995, Bedbrook and Bogorad 1976, Deng *et al.* 1989, Herrmann *et al.* 1975, Kowallik and Herrmann 1972, Lilly *et al.* 2001). The structure of the plastome in the green lineage outside of Streptophytina is significantly more divers (reviewed in Maier and Schmitz-Linneweber 2004). The circular-mapping plastid genome of land plants is characterised by the subdivision into four sections, where two inverted repeats, usually containing an rRNA operon, separate the LSC (Large Single Copy) from the SSC (Small Single Copy) region. The genes preserved in the plastome usually encode components of the genetic apparatus and parts of the photosynthetic apparatus, as well as some gene products involved in biosynthetic pathways localised in plastids (see Fig.2 and Tab. 2). Beside the highly conserved gene repertoire, even intron content and positions are stable in plastids of angiosperm plants. The strongest deviations from this conserved structure are found in parasitic plant species that have adopted a non-photosynthetic lifestyle. E.g., *Epifagus virginiana* bears one of the smallest plastomes of all land plants known to date, encoding only 42 genes, with photosynthesis related genes and genes for the plastid encoded polymerase (PEP) lost or degraded as well as a number of tRNA and ribosomal protein genes.

Nevertheless this deviation underlines a more general trend in the evolution of plastids. Other than the chondrom (see 1.2.1) the plastome is a streamlined, compact, largely genic and evolutionary conservative genome and disfavours the integration of futile foreign DNA (Maier and Schmitz-Linneweber 2004).

Table 2: Genes encoded by the *Physcomitrella patens* plastid genome

Gene products	Genes
Genetic apparatus	
RNAP	*rpoB, C1**, C2*
Translation factor	*infA*
Ribosomal proteins	*rps2, 3, 4, 7, 8, 11, 12**,#, 14, 15, 18, 19*
	*rpl2**, 14, 16**, 20, 21, 22, 23, 32, 33, 36*
Ribosomal RNAs	*rrn16*§, 23*§, 4.5*§, 5*§
Transfer RNAs	*trnA(UGC)**,§, C(GCA), D(GUC), E(UUC), F(GAA), G(GCC), G(UCC)**, H(GUG), I(CAU),*
	*I(GAU)**,§, K(UUU)**, L(CAA), L(UAA)**, L(UAG), fM(CAU), M(CAU), N(GUU)*§, P(UGG),*
	Q(UUG), R(ACG)§, R(CCG), R(UCU), S(GCU), S(GGA), S(UGA), T(GGU), T(UGU),*
	V(GAC)§, V(UAC)**, W(CCA), Y(GUA)*
Intron maturase like	*matK*
Photosynthetic apparatus	
Photosystem I	*psaA, B, C, I, J, M*
Photosystem II	*psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z*
Cytochrome b6/f	*petA, B**, D**, G, L, N*
ATP synthase	*atpA, B, E, F**, H, I*
Chlorophyll biosynthesis	*chlB, L, N*
Rubisco	*rbcL*
NADH oxidoreductase	*ndhA**, B**, C, D, E, F, G, H, I, J, K*
other	
Acetyl CoA carboxylase subunit	*accD*
Clp protease subunit	*clpP****
Unknown function	*ycf1, 2, 3***, 4, 10, 12, 66**

*Intron-containing gene, ** two introns, #transspliced exons, §IR-encoded, remarkably ccsA and rpoA are not plastom-encoded

1.2.3 Mitochondrial promoters

In fungal and plant mitochondrial genomes multiple promoters are facilitated (Kühn *et al.* 2005, Notsu *et al.* 2002, Sugiyama *et al.* 2005, Unseld *et al.* 1997).

A variety of mitochondrial promoters from several flowering plants have been described, while nothing is known about promoter usage in deeper branching lineages of land plants. In mono- and dicotyledonous plants sequence elements from 17 up to 26 nucleotides extending a few nucleotides beyond the transcription initiation site were found to be necessary for efficient and correct promoter usage *in vitro*. The majority of flowering plant promoters identified to date reveal the existence of a conserved 5'-CRTA-3' core promoter element upstream and in close proximity usually two to four nucleotides from the initiation site (see Tab. 3).

Table 3: List of Monocot and Dicot Mitochondrial Transcription Initiation Sites.

	Motif	Plant	Gene	Position	-1 1	Ref.
Monocots	CRTA	Os	*trnfM*	-235	CACTTCAGAAAAT**CGTA**TA AAAATCAAGCAAGAA	A
			rrn26	-221	TTGCTGAAAAAAT**CGTA**TA AAAATCAAGCAAGAA	A
			rps3	-320	TTCAAGTATTT**CGTA**TTGT AACAATATTCGATCG	A
			nad3	-423	CAATCAAATTAT**CGTA**AAT AGATAGTACGGTTGC	A
			orf483	-363	CAATGAGAAAATG**CATA**GA TAATGAGCTGCGCCC	A
			rrn18	-221	TGTTTTGAATTGA**CATA**GA GAAATCTTCCCACTA	A
			nad9	-1407	CTGTACTCT**CGTA**AGATGA TACACGCTAGCTAAA	A
			nad9	-1329	GTAACCATATA**CATA**T AAAATGACAGCAA	A
			atpI	-219	GCAATTCATTCG**CATA**AAT GAAAAAAAAAGCGAG	A
			trnH	-448	CAAGCGGATTATA**CGTA**TG TAAACAGCCTTCCCG	A
		Sb	*atp6-2*	-188	CTCAGAAAA**CGTA**TAAT ATAGTCTCATTGGCC	A
			orf265	-547	GCACGAATTAACT**CGTA**CA TAACATAAAAAGTGA	A
			orf130	-116	TCTAAAATTAT**CATA**G AGAAAGATGTTCTGA	A
		Ta	*atpA*	-342	CGATTTCGAAATAG**CGTA**A AGTGATTCTTGCAAT	A
			coxII	-170	CAGAAAACG**CGTA**TAGT AAGTAGTCT	A
			orf25	-170	CGTCGAAAACG**CGTA**TAGT AAGTAGTCTTCGTCG	A
			coxIII	-323	ATGAAATTTTG**CATA**GAAT GATTGTCATGTTCAC	A
		Zm	*rrn26*	-185	GAAAAGAAAA**CGTA**TAAA AATCAAGCAAGAGGA	A
			rrn18	-232	TTTTGAATTGA**CATA**GATA AATCTTTATCGCGTT	A
			coxIII	-315	GATGAGAATTGA**CGTA**TTC AAAGTGGAAAGAACG	A
			coxIII	-355	TGAGAAATGA**CGTA**TCTTA CGTATCGAATCTCAA	A
			atp6	-624	TCTCCCATTTT**CATA**GAGA AAGATGTTCTGATTC	A
		Zp	*coxII*	-907	CTACGAAAACT**CGTA**TAAT AGTCTCATTGGCCAA	A
	YYTA	Os	*nad9*	-1029	AAGAAGATATG**TCTA**AAGT AGATGGTGATTGATA	A
			atpI	-470	TGGATAAATTAC**TTTA**GAA TAAGAAGTTCATGTT	A
			atpI	-273	TCAAGTTTTAAAT**TCTA**GA GAATAGCTCCGATCA	A
		Sb	*atp9*	-318	TCCTGTCCCCG**CCTA**ATCA CGTCAGATACTTGTA	A
			alp6-1	-442	ACGTGAATAAA**CTTA**CAAT GGAGATCTACTGTGG	A
			orf107	-536	ATTATTCAGATTAT**CTTA**A ATAATGGTCTGAGTA	A
		Zm	*atp9*	-288	ATGATCGGGTTA**TTTA**TTC AAATTGGAAATAGGC	A
			atp9	-496	GGCGTGTAACCA**TTTA**ACA GTTCAAAACCCAACA	A
	DDTA	Os	*nad9*	-486	AAGATCATTTGG**TGTA**GTT GAAATAACGGATTGG	A
			atpI	-678	CTTCCCTTCTAT**TGTA**GAA AGAGTTCTCAGCCAT	A
			nad9	-422	TATAAGAATG**AATA**ACTTT CTCAATGGTTTATTT	A
			atp1	-441	ATGTTTTATAGACT**AGTA**A GATAACTAACTAACT	A
		Sb	*orf107*	-358	CAATCCCAGACAAA**TATA**T TAATAAAGCATCCTT	A
			orf107	-295	CAGGCAAAAAAATC**AATA**G TAGATAATGAAAGAA	A
		Zm	*atp9*	-271	TTCAAATTGGA**AATA**GCTA GAACTCCTGCGCAGG	A
			atp9	-418	CGGTACTACTAGGT**TATA**G CTTAACCGACCCAGG	A
			atp1	-2300	TTAAGCCGAAA**AGTA**ACGT ATTAAAAAACAAAG	A
	NC	Os	*atp1*	-427	AGTAAGATAACTAACTAGT TAAGTAATACGAATC	A
			nad9	-1182	CATTATCTTGATCTCTCCT CCTTTGTATAAATAT	A
		Zm	*atp9*	-250	ACTCCTGCGCAGGCTTGCT GTACTGAATCTCTTT	A
			atp9	-218	TATAGTCTACCTGTCAAGC CCAAGATAAGGGCAG	A
		Zp	*coxII*	-349	CCGAACGTGTGCATGGAGA AAGCGATAATATTCG	A
Dicots	CRTA	Gm	*atp9*	-521	AAGACGAAATAG**CGTA**AGA GAAGAAAGTAGCAAG	A
			rna a	-526	AAGACGAAATAG**CGTA**AGA GAAGAAAGTAGAAAG	A
			rna b/c		ACTATAAAATTT**CATA**AGA GAAGAAAGCTGCTAG	A
		N s	*orf87*	-213	CACTCGAAATAT**CATA**AGA GAAGAAAAGCAAGTT	A
		Ob	*trnF*	-292	ACGACTTAATTT**CATA**AGA GAGAGAAGTGCCGGA	A
			cox2	-207	AACTAAAAATCT**CGTA**TGA GAATCAAAAGAATCT	A
			atpA	-209	AAGTTGATAAAT**CATA**AGA GAAGCAAAGTCCCTA	A
			rrn18	-125	AGATTGAAATGT**CATA**AGT GATGTTCGAAATCGC	A
		Ps	*atp9*	-1048	ACGAAATAATAG**CATA**AGA GAAGATATTGGACAA	A
			rrn18	-462	CATCACAAATAT**CGTA**GGT GATATCCGAAACCGC	A
		St	*rrn18*	-69	CTTAAGCAAAAT**CATA**ACA GAAGAAAAGTTCACA	A
			trnN1	-2087	TCAATCAAATGG**CATA**GTA TTAGGGTAGGAGGCT	A
		At	*ccmB*	-140	GTCTTTTAGTAAG**CGTA**TAT AAGCAGCAGTTTAGT	C
			rrn18	-156	TAGAATAATA**CGTA**TAT AATCAGAA	B
			orf291	-307	TGGAATAATA**CGTA**TAT AATCAGAT	B
			atp6-1	-200	GCCAATAATA**CGTA**TAT AAGAAGAG	B
			atp9	-295	CTGGTGCTCT**CGTA**TAT AAGAGAAG	B
			atp1	-1947	CTGGTGGTAT**CGTA**TAT AAGAGAGA	B
			cox2	-210	ATGTTGGTTT**CGTA**TAT AAGAAGAC	B
			tRNA-fMet	-98	TTTGAAATAT**CGTA**AGA GAAGAAGG	B
			rrn26	-893	CTATCAATTT**CATA**AGA GAAGAAAG	B
			atp1	-1898	CTATCAATTT**CATA**AGA GAAGAAAG	B
			atp9	-239	CTATCAATTT**CATA**AGA GAAGACGA	B
			atp6-1	-156	CTATCAATCT**CATA**AGA GAAGAAAT	B
			atp6-2	-148	CTATCAATCT**CATA**AGA GAAGAAAT	B
			atp8	-157	CTATCAATCT**CATA**AGA GAAGAAAT	B
			atp8	-228/226	CATACCATAA**CATA**TAT AGAATCGA	B
	DDTA	At	*rrn18*	-424	TCAAATCCTC**GGTA**TAT AAAGAGAA	B
			cox2	-683	GACACGTAA**GGTA**AAAT AAGAATCT	B
			rps3	-1053	TTTTTTATTT**GGTA**GGT AACATCGC	B
			atp9	-487	ATGTCTTATT**GGTA**TGT GATACAAG	B
			atp9	-652	AGAAGATTGA**AGTA**AGG AGCAGGTT	B
			atp6-2	-436	TCTTGAATTA**AGTA**TAT AGAAAAGA	B
			atp6-2	-507	GATAAATTA**AGTA**TAGT AATAAGAA	B
			rrn18	-69	AAAGTGGAATTG**AATA**AGA GAAGAAAGGTCCACA	A
			atp8	-710	ATCGGAGCTGCC**AATA**A GCTAATCC	B
			atp8	-999	ATAAAATTA**AATA**AAGA GCAAAAAT	B
			rrn18	-353	TACTTTTCCATC**TATA**T AAAATGAA	B
			atp6-1	-913	AGCCCTTTATAT**TATA**TAAT AAAGC	B
			atp6-1	-916	AGCCCTTTAT**ATTA**TAT AATAAAGC	B
			cox1	-355	AATTTATTCA**ATTA**TAT AATAATAA	B
			cox2	-481	ATGAATATTC**ATTA**GAT AATAGATT	B
			rps3	-1133	TAGAAAAAATT**ATTA**GT AATACGTA	B
			tRNA-fMet	-574/573	CTAATTTA**TATA**AAAA AGACCGGGA	B
	NC	Gm	*rnae*		TTTTTTGTTTGGATTGATC GAGTGTACTTTTTCT	A
		Ob	*cox1*	-1473	TCCTGTTCAACAATTGCGT AAGTGAGGGTGGAGG	A
			coxIII	-337	TCCTGTTCAACAATTGCGT AAGTGAGGGTGGAGG	A
		St	*atp9*	-128	TGTGAAGTCTACCGCCTGT CTAGCCTATGCTTTG	A
			rrn26	+1	GGGTACAAGATCGAAAAGA ATGCATTGGATGGAT	A
			trnS	-306	AATGTAATAGAAGAGGCAG AGTCCAGTTCCCCTC	A
			trnN2	-135	GTCCGGAGCAGGAGCATTA GGGTAGGCGATGGCC	A
			nad2	-454	TTCGTCAGCGTAGGAAGGC CAACCGAGCGAAGCT	A

Abbreviations are as follows: Os: *Oryza sativa*; Sb: *Sorghum bicolor*; Ta: *Triticum aestivum*; Zm: *Zea mays*; Zp: *Zea perinnis*; At: *Arabidopsis thaliana*; Gm: *Glycine max*; Ns: *Nicotiana sylvestris*; b: *Oenothera berteriana*; Ps:

Pisum sativum; St: *Solanum tuberosum*; (A) (Fey and Marechal-Drouard 1999); (B) (Kühn *et al.* 2005); (C) (Forner *et al.* 2007)

For dicot mitochondrial promoters two classes of nonanucleotide consensuses, 5'-CRTAAGAGA-3' and 5'-CGTATATAA-3' are depicted (Binder *et al.* 1996, Fey and Marechal-Drouard 1999, Kühn *et al.* 2005). However, almost as many promoters deviate not only in sequences surrounding the transcriptional initiation site, but also in the 5'-CRTA-3' core element. The monocot promoters identified to date are more divers and for some a 5'-YYTA-3' core element, not found in dicot promoters is described. Some mitochondrial genes of flowering plants are transcribed from non-consensus promoters showing no conservation at all in the sequence surrounding the transcription start site (TSS). A common feature of both monocots and dicots is that initiation of transcription is preferred at adenine. In dicots also guanine serves frequently as the initiating nucleotide and purines in general appear to be obligatory at positions +1 and +2 with respect to the TSS while only moderately frequent at positions +3 to +8. Again this pattern is much less pronounced in promoters of monocotyledonous plants. Sequences upstream from the TSS are usually A/T-rich. This A/T-richness has been described to be important for proper promoter function in both monocots and dicots (Dombrowski *et al.* 1999, Rapp *et al.* 1993, Rapp and Stern 1992) and is a structural feature of mitochondrial promoters in fungi (Tracy and Stern 1995). Interestingly, the yeast consensus motif is a nonanucleotide with the well conserved consensus sequence 5'-ATATAAGTA-3' encompassing the transcriptional initiation site (+1) and the eight nucleotides directly upstream. Especially considering 5'-DDTA-3' like promoters this element appears to be neither significantly different from plant promoter elements, nor with respect to the initiation site. Furthermore, promoter strength in yeast is determined by the nucleotide in position +2 (Biswas 1999, Biswas *et al.* 1985, Biswas and Getz 1986a, Biswas and Getz 1986b, Biswas *et al.* 1987). A purine (A or G) is found in highly efficient promoters, whereas weaker promoters contain a pyrimidine (C or T) in position +2 (Biswas and Getz 1986b, Wettstein-Edwards *et al.* 1986). Whether plant promoters with purines in position +1 and +2 show elevated promoter strength *in vivo* is unclear.

In flowering plants a mitochondrial gene is often transcribed from multiple promoters. Lupold et al. suggested that this might be a mere reflection of a promiscuous transcriptional apparatus or indicate that multiple promoters are needed to counteract possible consequences of frequent intramolecular recombinations of the mitochondrial genome (Lupold *et al.* 1999).

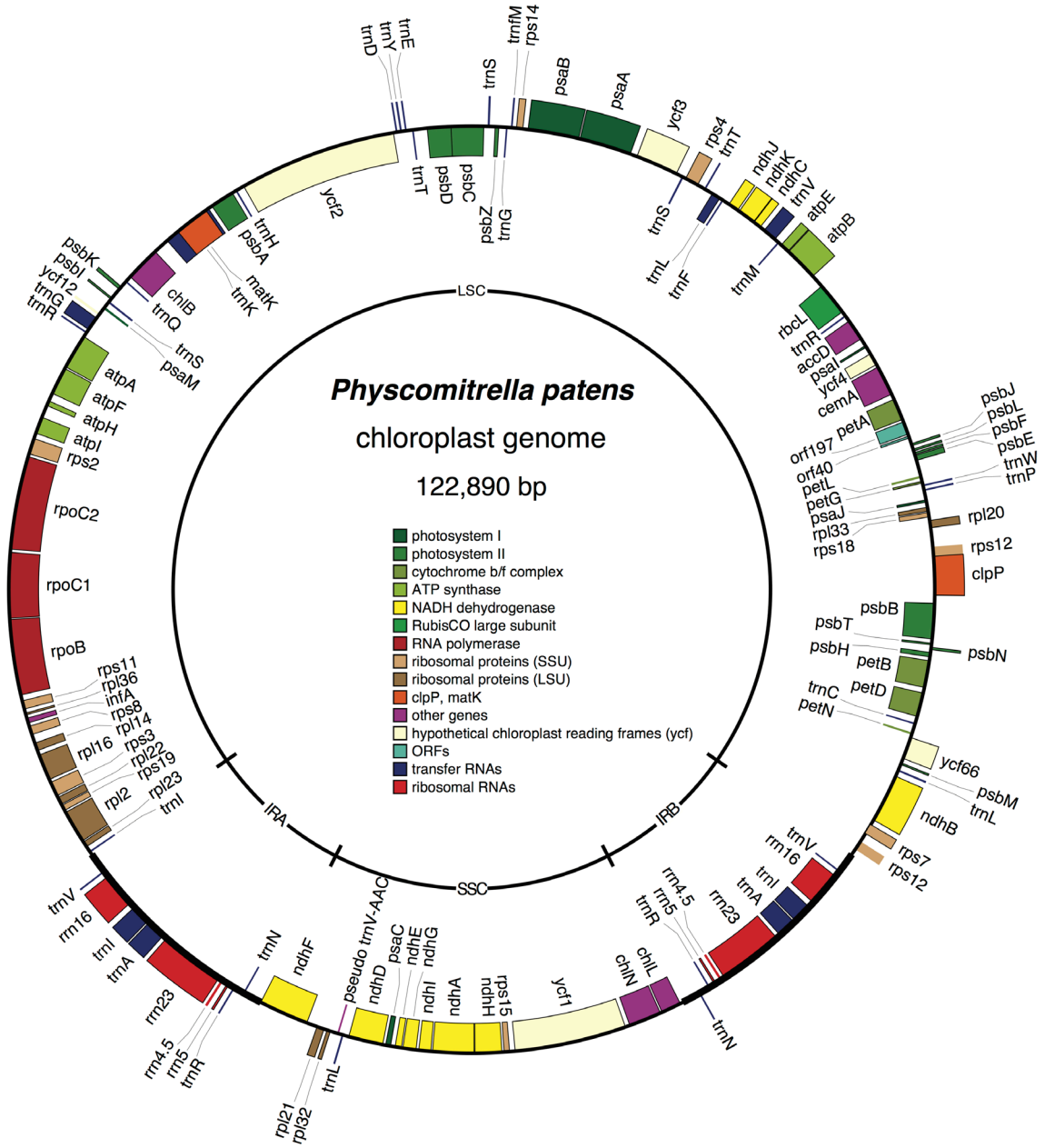


Figure 2: Circular map of the plastid genome of Physcomitrella patens. Genes encoded by the counter clockwise strand are shown and indicated outside the circle, whereas genes encoded by the clockwise strand are inside. The map was created by OGDRAW (Lohse *et al.* 2007) and embellished in Adobe Illustrator version CS3.

1.2.4 Plastid promoters

Transcription of a subset of tRNA genes from spinach (Cheng *et al.* 1997, Gruissem *et al.* 1986) mustard (Liere and Link 1994, Neuhaus and Link 1990, Nickelsen and Link 1990), and *Chlamydomonas* (Jahn 1992), is assumed to be driven by internal promoters. The transcription start site of the spinach trnS gene is located 12 nucleotides upstream of the mature tRNA coding region (Wu *et al.* 1997) and while the coding region (+1/+93) can promote basal levels of transcription *in vitro*, introduction of an AT-rich region directly upstream (-31/-11), was shown to restore wild-type promoter strength. Furthermore, the *trnS* coding region contains sequences analogous to the A and B blocks of nuclear tRNA promoters

which are transcribed by the eukaryotic RNA polymerase III (Galli *et al.* 1981, Geiduschek and Kassavetis 1995). Efficient transcription of *trnR* from *Pelargonium* was observed after injection into *Xenopus* oocytes (Hellmund *et al.* 1984), in turn suggesting that the plastidial tRNAs might be transcribed by an RNA polymerase III-type enzyme in plastids. Nevertheless, no such enzyme was detected in plastids of any plant investigated thus far.

Instead, transcription of plastid genes was found to be dependent on two other strikingly different activities, one specified as a nuclear encoded transcription activity and the other represented by the plastid encoded bacterial-type polymerase (PEP) inherited from the cyanobacterial endosymbiont (reviewed in Hess and Börner 1999, Liere and Börner 2007). Accordingly, a subset of plastid promoters contains the typical eubacterial-like promoter architecture, with variations of this theme such as additional *cis*-elements and/or differential conservation of the -10/-35 regions (Liere and Börner 2006). In fact, some of those σ^{70}-like promoters were shown to be efficiently recognised by the *E. coli* polymerase holoenzyme (Boyer and Mullet 1988, Bradley and Gatenby 1985, Eisermann *et al.* 1990, Gatenby *et al.* 1981). In Figure 1 different PEP promoter compositions are depicted. For some PEP promoters containing an additional TATATA element directly upstream from the -10 region it was demonstrated that transcription is efficiently initiated *in vitro* without a -35 element. Importance of this finding is further substantiated by the observation of some putative plant PEP promoters lacking a conserved -35 region all together (personal communication K. Liere). Utilisation of the *cis*-elements of the highly conserved *psbA* promoter from barley, mustard and wheat is not only different between those plant species, but also species specific deviations are observed dependent on the developmental stage and between dark- and light-grown plants, suggesting differential usage of additional *trans*-factors (reviewed in Liere and Börner 2006). Whereas in the upstream regions of photosynthetic genes and operons PEP promoters are located, genes for housekeeping functions such as transcription and translation contain PEP as well as NEP promoter elements upstream of the coding regions (see Fig.3). Genes exclusively transcribed from NEP promoters found to date are rare, including *accD*, *clpP* and *ycf2* from dicots, *rpl23* from monocots and the *rpoB* operon from all flowering plants investigated thus far.

Due to their low abundance NEP primary transcripts were mapped mostly in PEP deficient plants. Identification of conserved upstream sequences led to the proposition of three types of NEP promoters designated Type Ia, Type Ib and Type II as shown in Figure 3 (Hess and Börner 1999, Liere and Maliga 2001).

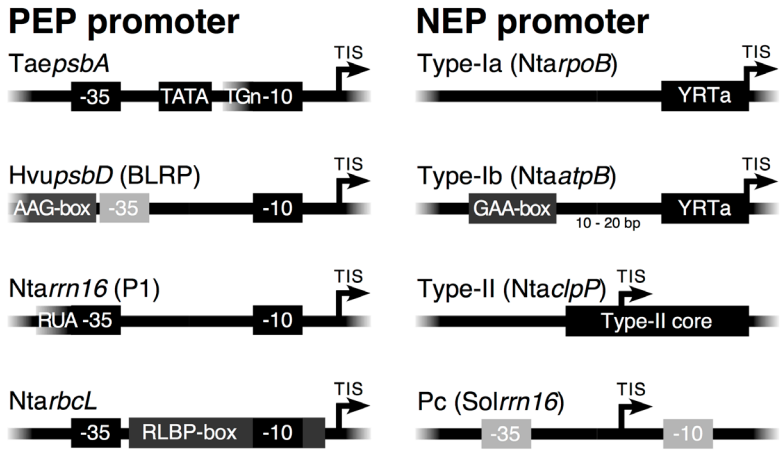


Figure 3: Overview of different types of PEP and NEP promoters (taken from Liere and Börner 2007). PEP promoter: the wheat *psbA* (Tae*psbA*), barley *psbD* BLRP (Hvu*psbD*), tobacco *rrn16* (Nta*rrn16*), and the tobacco *rbcL* PEP promoters (Nta*rbcL*) are shown. Conserved -10/-35 consensus elements, as well as individual promoter elements as the TATA-box , extended -10 sequence (TGn), AAG-box, RUA-element, and RLBP-binding region are indicated (RLBP, *rbcL* promoter-binding protein). The less conserved -35 element in the barley *psbD* blue-light- responsive promoter (BLRP) is shown in grey. NEP promoter: typical architectures of Type-I, Type-II, and Pc NEP promoters from tobacco and spinach are shown (names in brackets). The YRTa promoter core and GAA-box are marked; TSS: transcription initiation site, indicated by arrows. The -35 and -10 elements not used in spinach *rrn16* Pc promoter recognition are shown in grey.

Type I promoters resemble the structure described for mitochondrial promoters. A 5'-YRTA-3' is located 2 to 5 nucleotides upstream from the site of transcription initiation whereas in type Ib promoters an additional conserved sequence motif is found approximately 18-20 bp upstream of the 5'-YRTA-3' core element designated as box II or GAA box (Kapoor and Sugiura 1999, Silhavy and Maliga 1998). Functional relevance of the GAA box was shown for the tobacco *atpB*-289 promoter *in vitro* as well as *in vivo* (Kapoor and Sugiura 1999, Xie and Allison 2002). While most described NEP promoters contain an 5'-YRTA-3' core element (type-I NEP promoters), several others lack such motif and therefore are described as type II NEP promoters, again resembling the situation found in mitochondria of flowering plants, where such upstream sequences are described as nc (non-consensus) promoters (see 1.2.4). In *Arabidopsis* this class of type II NEP promoters is represented most prominently by P*clpP*-58 and the Pc P*rrn16*-139 promoter (Swiatecka-Hagenbruch *et al.* 2007). The *clpP* promoter sequence seems to be conserved from liverwort to flowering plants, but seems to be transcriptionally silent in monocotyledonous plants like rice (Sriraman *et al.* 1998a). However, transgenic tobacco plastids carrying the rice *clpP* promoter sequence allow transcription initiation, suggesting the need for type II specific trans factors or usage of a second NEP activity not present in monocots (Sriraman *et al.* 1998a). An exceptional promoter is represented by the *rrn* PC promoter shown to be recognised by NEP2 with the help of CDF2-B in spinach (Bligny *et al.* 2000, Iratni *et al.* 1994, Iratni *et al.* 1997). This promoter is used in spinach, *Arabidopsis*, and mustard (Baeza *et al.* 1991, Pfannschmidt and Link 1997, Sriraman *et al.* 1998b). In other eudicotyledons like tobacco, carrot, pea, and in all other thus far analysed monocotyledons, transcription initiation at PC could not be demonstrated, and transcription of the *rrn* operon instead is initiated at the eubacterial type P2 promoter (for review see Lerbs-Mache 2000 and Liere and Börner 2007).

In general, almost all chloroplast genes are transcribed from several independent NEP and/or PEP promoters (Swiatecka-Hagenbruch *et al.* 2007) and both mono- and polycistronic transcripts are found (Barkan 1988, Haley and Bogorad 1990, Kahlau and Bock 2008). This is further substantiated by the finding, that in PEP deficient plants all transcripts though some seem to be reduced are detectable (Legen *et al.* 2002). The polycistronic nature of some primary transcripts might be a mere result of relaxed transcription termination of the NEP activity but also and more importantly an inherited feature with respect to the eubacterial origin of the plastome and its PEP activity. Almost the same situation is observed in 'higher plant' mitochondria (Dombrowski *et al.* 1997), where downstream genes are cotranscribed with upstream partners (Schuster 1993) even in cases where they owe their own proximal promoter, introducing the need for posttranscriptional processing mechanisms in both plant organelles.

1.2.5 Transcriptional activities in plant organelles

In spite of their eubacterial origin, plastids and mitochondria seem to apply different strategies to transcribe the genomes they contain. In concordance with their endosymbiotic origin most plastid genomes encode all subunits for a multicomponent eubacterial-like core polymerase (PEP), showing highest similarity with cyanobacterial α, β, β′ and β′′ subunits. Additional cofactors such as nuclear encoded sigma factors are facilitated in a promoter specific or light dependent manner and are suggested to be contingent on the developmental stage (reviewed in (Liere and Börner 2006)). Although the PEP has an essential role in transcription (see 1.2.5) a second nuclear encoded polymerase (NEP) activity is present in plastids (Hess and Börner 1999, Liere and Börner 2007, Liere and Maliga 2001, Link 1996).

Early on it was suggested that the NEP activity might be represented by a single polypeptide, phage-type polymerase (Lerbs-Mache 1993). Candidate NEP genes emerged with the identification of three *Arabidopsis* genes, designated *RpoTm*, *RpoTp* and *RpoTmp* encoding different but related phage-type RNA polymerases. Localisation studies of deduced N-terminal transit peptides suggest that RpoTp and RpoTm are imported into plastids and mitochondria, respectively. The N-terminus of a third enzyme of this type, RpoTmp, mediates dual targeting into both organelles (Hedtke *et al.* 1997, Hedtke *et al.* 2000, Hedtke *et al.* 1999a). Therefore, in *Arabidopsis* both plastids and mitochondria very likely contain two phage-type transcriptional activities each. Whereas in plastids the eubacterial transcriptional machinery is still found in all photosynthetically active plants, mitochondrial transcription appears to be strictly dependent on phage-type DNA dependent RNA polymerases since, all eukaryotes with the exception of jacobids including *Reclinomonas americana*, seem to lack a mitochondrial bacterial-type RNA polymerase (Gray *et al.* 2004, Lang *et al.* 1997).

An exceptional situation is observed in plastids of parasitic angiosperms where a gradual loss of coding capacity of photosynthetic genes is accompanied by a switch from integrated NEP and PEP to purely NEP based transcription, with some parasites showing complete loss of PEP genes (Krause 2008). *RpoT* genes encoding mitochondrial or plastid phage-type polymerases have been identified in other land plants. Whereas eudicotylodonous plants like *Arabidopsis* seem to posses at least three *RpoT* genes (Hedtke *et al.* 2000, Hedtke *et al.* 2002, Kobayashi *et al.* 2002, Kobayashi *et al.* 2001a, Kobayashi *et al.* 2001b), monocots investigated thus far harbour not more than two, one coding for a mitochondrial RNA polymerase and the other for a plastidial enzyme (Chang *et al.* 1999, Ikeda and Gray 1999b), seemingly lacking the gene encoding the dual targeted RpoT polymerase found in eudicots. Dual targeting of *RpoT* gene products to both mitochondria and plastids being the result of

two different translational starts has been reported for two phage-type RNA polymerase encoding genes of *Physcomitrella patens* (Richter *et al.* 2002). For both PpRpoTmp1 and PpRpoTmp2, translation initiation at the first of two in-frame AUG start codons was found to compose a polypeptide with N-terminal plastidial targeting properties, whereas initiation at the downstream AUG gave rise to a mitochondrial protein (Richter *et al.* 2002). More recently Kabeya and colleagues were able to show that translation initiation on both *RpoTmp1* and *RpoTmp2* transcripts is likely regulated, since initiation of translation from the first AUG was repressed for all constructs with the native 5'UTR present. Similar results are reported for the *Arabidopsis* polymerase RpoTmp, clarifying the need to further investigate the spatio-temporal localisation of RpoT proteins with dual targeting properties *in planta*. Thorough expression studies reported by Emanuel and colleagues reveal that in *Arabidopsis* all three *RpoT* genes are actively transcribed in roots, leaves, and flowers and whereas *RpoTp* is most expressed in green tissues, *RpoTm* and *RpoTmp* expression appears to be most active in meristems and young cells. Furthermore, overlapping spatio-temporal expression patterns of *RpoTm* and *RpoTmp* are found, whereas *RpoTp* expression differs significantly from both. Therefore, the authors suggested that RpoTm and RpoTmp while expressed in the same tissue might facilitate differential expression of the mitochondrial genome but also that the RpoTp enzyme exclusively found in plastids might play a major role in chloroplast transcription, while the plastidial portion of RpoTmp would have its main function in the transcription of genes in non-green tissues (Emanuel *et al.* 2006). In rice expression analysis is indicative of NEP activation occurring at a strictly limited stage of leaf development, preceding the expression of PEP and of the photosynthetic apparatus (Kusumi *et al.* 2004). In developing plastids of maize transcription of RNAs with NEP promoters was shown to increase but stability of the transcripts decreased. As a result relatively stable transcript accumulation was observed for NEP genes. For PEP transcribed RNAs however increased transcription and stability resulted in elevated transcript accumulation in older leaf sections. Both findings together led the authors to suggest that in maize plastids two mRNA stability classes exist correlating with NEP activity on one side and PEP activity on the other (Cahoon *et al.* 2004).

First direct evidence that transcription in plant organelles depends on nuclear encoded phage-type RNA polymerases or concerning plastids: that NEP activity is represented by RpoT enzymes was provided by Liere and colleagues (Liere *et al.* 2004). Analysis of plastid transcription in *RpoTp* overexpressing tobacco plants facilitating the *atpB* locus which contains both NEP and PEP promoter elements revealed that initiation from a typical type I NEP promoter was enhanced (Liere *et al.* 2004). More recently additional evidence was presented by Swiatecka et al., showing that the NEP promoter of *ycf1* was not used in *Arabidopsis* mutant seedlings lacking RpoTp activity. Furthermore, they provided evidence for NEP being represented by two phage-type RNA polymerases (RpoTp and RpoTmp) that have overlapping as well as gene-specific functions in the transcription of plastidial genes, since they found almost all NEP promoters to be used in plants with low or lacking RpoTp activity (Swiatecka-Hagenbruch *et al.* 2008). A more specific function for RpoTmp in plastids of *Arabidopsis* was recently proposed by Courtois et al. in transcribing the *rrn* operon from the PC promoter (see 1.2.5) during seed imbibition, while RpoTp was shown to be the major NEP activity recognizing type 1 and type 2 NEP promoters (Courtois *et al.* 2007).

Even though very likely, since almost all eukaryotes facilitate phage-type RNA polymerases for mitochondrial transcription and targeting studies indicate the presence of such proteins in plant mitochondria as stated above, direct evidence for phage type RNA polymerases being the transcriptional activity in plant mitochondria was still lacking.

Other than monocotyledonous plants, mosses and eudicots encode at least two phage-type RNA polymerases targeted to mitochondria. The relative importance of more than one putative transcriptional activity in mitochondria is completely unclear. At present there are no data directly indicating a preference of those polymerases for distinct sets of promoters *in vivo* (see 1.2.4) even though RpoTm but not RpoTmp has been shown to specifically recognize multiple mitochondrial promoters *in vitro* (Kühn *et al.* 2007). Differential expression in certain tissues and/or developmental stages is rather unlikely since *RpoTm* and *RpoTmp* genes in *Arabidopsis* are reported to display overlapping expression patterns in different tissues and at different developmental stages (Emanuel *et al.* 2006). Comparison of plastidial and mitochondrial promoters in *Arabidopsis* revealed no common motifs that might direct transcription initiation by RpoTmp (Kühn *et al.* 2005, Swiatecka-Hagenbruch *et al.* 2007). Therefore the transcriptional role of AtRpoTmp in mitochondria is thus completely unclear from the current literature and even less is known about the transcriptional machinery in mitochondria of other plant species like *Physcomitrella patens* or *Nicotiana tabacum*.

In general mitochondrial transcription in higher plants is reported to be non-stringently controlled (Holec *et al.* 2008b). Transcription initiation from multiple sites not confined to the upstream regions of annotated mitochondrial genes, may result in presumably non-functional RNAs and occasionally even detrimental transcripts usually efficiently removed by mitochondrial RNA-degrading mechanisms (Finnegan and Brown 1990, Holec *et al.* 2006). Post-transcriptional processes have been identified as a major regulator of mitochondrial RNA pools, while the abundance of an individual transcript does not always reflect transcript synthesis rates (Giegé *et al.* 2000). A major paradigm of plant mitochondrial gene expression is that transcriptional mechanisms in mitochondria are not significant for the regulation of mitochondrial function, but regulated by posttranscriptional, translational and posttranslational processes (Binder and Brennicke 2003, Gagliardi and Gualberto 2004, Giegé and Brennicke 2001). It is therefore surprising that in the mitochondria of eudicotyledonous plants and mosses the basic process of transcription involves more than one RNA polymerase.

1.2.6 Cofactors of phage-type RNA polymerases in plant organelles

Similarity of plastid NEP promoters with plant mitochondrial promoter elements most likely recognised by RpoT enzymes raises the question whether similar transcriptional cofactors interact with these core RNA polymerases.

Thus far the DNA binding protein CDF2, isolated from spinach chloroplasts is the only known example of a factor stimulating transcription from a NEP promoter (Bligny *et al.* 2000). However, structural details of CDF2 have not been shown to date, and no CDF2 like activity has been purified from plant mitochondria. For *Arabidopsis* Sig2 and its maize orthologue Sig2B dual localisation in plastids and mitochondria was implicated (Beardslee *et al.* 2002, Tandara 2000). Since PEP subunits are absent from mitochondria Sig2 might be a potential candidate to be involved in phage-type polymerase transcription. On the other hand experimental data available to date relate the function of Sig2 merely to the bacterial-type PEP activity in plastids (Beardslee *et al.* 2002, Kanamaru and Tanaka 2004). A regulatory role was recently suggested for the plastid-encoded $tRNA^{Glu}$ in chloroplast transcription (Hanaoka *et al.* 2005). From gel retardation experiments the authors concluded that $tRNA^{Glu}$ has a specific affinity for RpoTp. Moreover, transcription from a plastidial *accD* promoter, which is considered to be catalyzed by a phage-type RNA polymerase, was inhibited by the addition of $tRNA^{Glu}$ to transcriptionally active proplastid extracts from *Arabidopsis*. This led the authors to concluded that $tRNA^{Glu}$ might mediate a switch in RNA polymerase utilisation from NEP to

the PEP during chloroplast development. To the contrary, in a recent publication it was shown that all tRNAs tested inhibited transcription of RpoT enzymes (Bohne *et al.* 2009). Moreover, the inhibiting molecular ratio of 100 tRNAs per 1 molecule polymerase might suggest a more general biochemical effect, possibly based on the polymerisation reaction kinetics.

Previous studies suggested that factors supporting mitochondrial RNAPs in promoter recognition and/or transcription initiation should exist also in higher plants (Binder and Brennicke 2003, Newton *et al.* 1995, Young and Lonsdale 1997). Furthermore, unlike the single-subunit RNA polymerases of bacteriophages, mitochondrial phage-type RNA polymerases in fungi and metazoans require auxiliary factors to initiate transcription at promoter sequences. Considering the homologous nature of the mitochondrial RNA polymerases in eukaryotes, one would assume that plants also need auxiliary factors for mitochondrial transcription, possibly similar to those found in yeast or human mitochondria. To date, two types of such nuclear encoded mitochondrial transcription factors, designated here as mtTFA and mtTFB, have been characterised in *S. cerevisiae* (Matsunaga and Jaehning 2004b, Schinkel *et al.* 1987, Winkley *et al.* 1985), *X. laevis* (Bogenhagen 1996, Bogenhagen and Insdorf 1988), *Drosophila melanogaster* (Matsushima *et al.* 2005, Matsushima *et al.* 2004), humans (Falkenberg *et al.* 2002, Fisher and Clayton 1988, McCulloch *et al.* 2002) and mouse (Gaspari *et al.* 2004). In human and yeast mitochondria the HMG-box protein mtTFA is an abundant DNA-binding protein that enhances transcriptional activity (Kanki *et al.* 2004).

mtTFBs belonging to the family of *S*-adenosyl-L-methionine (SAM)-dependent rRNA adenine dimethyltransferases were found to be essential for transcription in yeast and mammalian mitochondria (Asin-Cayuela and Gustafsson 2007, Jang and Jaehning 1991). Homologs of this rRNA adenine dimethylase protein family are found to function in all domains of life. (SAM)-dependent rRNA dimethylases such as KsgA from *E.coli* or the nucleolar Dim1 of *S. cerevisiae* usually mediate the name-giving methylation activity modifying two specific adenosines in a highly conserved stem-loop near the 3' end of ribosomal RNAs. Mitochondria of all metazoan species investigated thus far possess two mtTFB proteins. In contrast, mitochondria of *S. cerevisiae* and other fungal species possess only one mtTFB. Yeast sc-mtTFB lacks rRNA dimethyltransferase activity correlating with the lack of the corresponding modification in the mitochondrial rRNA of budding yeast (Cotney and Shadel 2006). Phylogenetic analyses indicate that fungal mtTFBs and metazoan TFB2M enzymes are paralogous to animal TFB1M proteins (Cotney and Shadel 2006), while all mtTFB homologs including animal TFB1M and TFB2M are likely descendants of the dimethyltransferase of the original mitochondrial endosymbiont (Shutt and Gray 2006b). Both human proteins, hs-TFB1M and hs-TFB2M, have been reported to bind SAM and to substitute for the *E. coli* KsgA in methylating the 16S rRNA at the correct sites (Cotney and Shadel 2006, McCulloch *et al.* 2002, Seidel-Rogol *et al.* 2003), though hs-TFB2M is a less efficient enzyme than hs-TFB1M (Cotney and Shadel 2006). Recently, Metodiev et al. (Metodiev *et al.* 2009) reported that mouse TFB1M is not acting as transcription factor in the presence of TFB2M but is a nonredundant dimethlytransferase. Disruption of the *Tfb1m* gene led to the loss of the 12S rRNA methylation and abolished mitochondrial ribosome assembly without affecting mitochondrial transcription.

The potential involvment of transcriptional cofactors in plant mitochondrial transcription was substantiated recently by Kühn et al. (Kühn *et al.* 2007). Since selective promoter recognition by RpoTm and RpoTp and an apparent inability of RpoTmp to correctly initiate transcription at mitochondrial and plastidial NEP promoters was observed, it was suggest that auxiliary factors are required for efficient initiation of transcription *in vivo* (Kühn *et al.* 2007). To date

no information is available about mitochondrial rRNA methylation in the green branch of life and so far mtTFA- and mtTFB- related transcription factor proteins have not been found in plants. Instead a potential transcription factor has been purified from wheat mitochondria. The 63-kDa PPR protein (p63) stimulates transcription initiation *in vitro* (Ikeda and Gray 1999a). Since PPR proteins have various functions in RNA metabolism (Lurin *et al.* 2004, Small and Peeters 2000), the unambiguous function of this protein as a transcription factor *in vivo* is still somewhat unclear, especially considering that an extract prepared from wheat mitochondria was able to accurately initiate transcription at the *cox2* promoter without the aid of additional p63 (Ikeda and Gray 1999a).

1.3 Origin and evolution of phage-type RNA polymerases

Other than the multi-subunit RNA polymerases from eukaryotes and bacteria, the RNA polymerase of the T-odd bacteriophages is a single-polypeptide enzyme that is solely responsible for sequence specific promoter recognition, for correct initiation of transcription and for elongation (Steitz 2004). Although eukaryotic phage-type RNA polymerases likely require auxiliary proteins to initiate transcription at organellar promoters (see I.3.3), the thoroughly studied T7 enzyme serves as a model for both bacteriophage and eukaryotic phage-type RNA polymerases. Alignments of phage and phage-type enzymes revealed conserved domains and motifs with identical or conservatively substituted amino acids (Chang *et al.* 1999, Hedtke 1998). Crystal structures of the T7 RNA polymerase have suggested the catalytically active C-terminal portion of the protein to fold into the “fingers”, “palm” and “thumb” subdomains (Jeruzalmi and Steitz 1998, McAllister and Raskin 1993, Sousa *et al.* 1993), structure elements typically found in the superfamily of nucleic acid polymerases, including DNA polymerases and reverse transcriptases (reviewed in Sousa 1996). Whereas highest sequence similarity scores of organellar RNA polymerases to the T7 enzyme are found in C-terminal portion including all the essential residues that participate in RNA polymerization (Hess and Börner 1999), the N-terminal regions diverge strongly. In the case of the T7 polymerase this region bears elements contributing to promoter recognition of the enzyme. For the yeast polymerase Rpo41 it has been reported that two regions the “specificity loop” and intercalating β-hairpin involved in promoter recognition and DNA melting by T7 RNA polymerase are conserved in structure, but not in sequence (Matsunaga and Jaehning 2004a, Matsunaga and Jaehning 2004b). This was substantiated by the finding that RPO41 is capable of accurate promoter-specific transcription initiation without sc-mtTFB *in vitro*. A recent study suggests that corresponding structures might be formed by plant RpoT polypeptides as well (Kühn 2006) and in turn could contribute to the observed intrinsic ability of RpoT enzymes to preferentially initiate transcription at particular sequences *in vitro* (Kühn *et al.* 2007). Initially, homologs of nuclear encoded single polypeptide RNA polymerases were found in T-odd phages and in a very heterogeneous group of linear mitochondrial plasmids (Cermakian *et al.* 1997). Even though clearly related, neither of the two seemed to be a likely source for the nuclear encoded single polypeptide polymerases, since the T-odd phages, a subgroup of the taxon *Podoviridae*, are lytic, γ-proteobacteria infecting viruses and the plasmid-encoded enzymes were only found in plant and fungal mitochondria (Cermakian *et al.* 1997). Therefore, the origin of nuclear encoded phage-type RNA polymerases and the timing of their acquisition remained enigmatic. The advent of bacterial and viral genomics permitted identification of prophage sequences encoding T7-like RNAPs in some eubacterial genomes outside from γ-proteobacteria. Most importantly,

considering the α-proteobacterial origin of mitochondria, a podoviral prophage sequence was identified in *Agrobacterium tumefaciens*. Therefore, it is clear now, that the host specificity of bacteriophages from the taxon *Podoviridae* is less stringent than thought before and that those phages are not always lytic (Filee and Forterre 2005, Shutt and Gray 2006a). Together these two findings add weight to the hypothesis of a podoviral origin of nuclear encoded phage-type RNA polymerases. Considering a subsequent gene transfer from mitochondria to the nucleus, a prophage that might have been part of the α-proteobacterial endosymbionts genome or an active podoviral phage, present in the α-proteobacterial ancestor of mitochondria at the time of the endosymbiotic uptake, is therefore a likely source for nuclear encoded phage-type RNA polymerases. Shutt and Gray (2006) seem to favour the latter variation of this hypothesis, because of the organismal distribution observed for the linear plasmids, which are proposed to have bacteriophage ancestry as well (Shutt and Gray 2006a).

While *RpoT* genes from angiosperms have been well characterised (reviewed in Hess and Börner 1999, Liere and Börner 2006, Liere and Börner 2007, Weihe 2004), almost nothing is known about the evolution of those genes in other lineages of the plant kingdom, with the only exception of two nuclear *RpoT* genes reported to exist in the leafy moss *Physcomitrella patens* (Kabeya *et al.* 2002, Richter *et al.* 2002). Phylogenetic analyses show a distinct group of the plant RpoT sequences clearly separated from those of fungi and animals. In concordance with this finding the plant specific gene structure is highly conserved in all plant *RpoT* genes investigated thus far (Richter *et al.* 2002). Within this group the RpoT polymerases of *Physcomitrella patens* are separated from the angiosperm polymerases, with the latter falling into two groups with plastid localised enzymes on one hand, and mitochondrial and dual-targeted polymerases on the other (Liere and Börner 2007, Richter 2003, Richter *et al.* 2002). The duplication event that led to the *RpoT* gene copies in *P. patens*, thus, must be considered to have occurred independently from the one giving rise to the *RpoT* gene family in angiosperms. Furthermore, it shows that the duplication giving rise to a group of exclusively plastid localised RpoTp gene products of angiosperms, is predating the duplication event that led to the appearance of a third *RpoT* gene as found in eudicot plants, such as *A. thaliana* or *N. sylvestris* (Richter 2003, Richter *et al.* 2002).

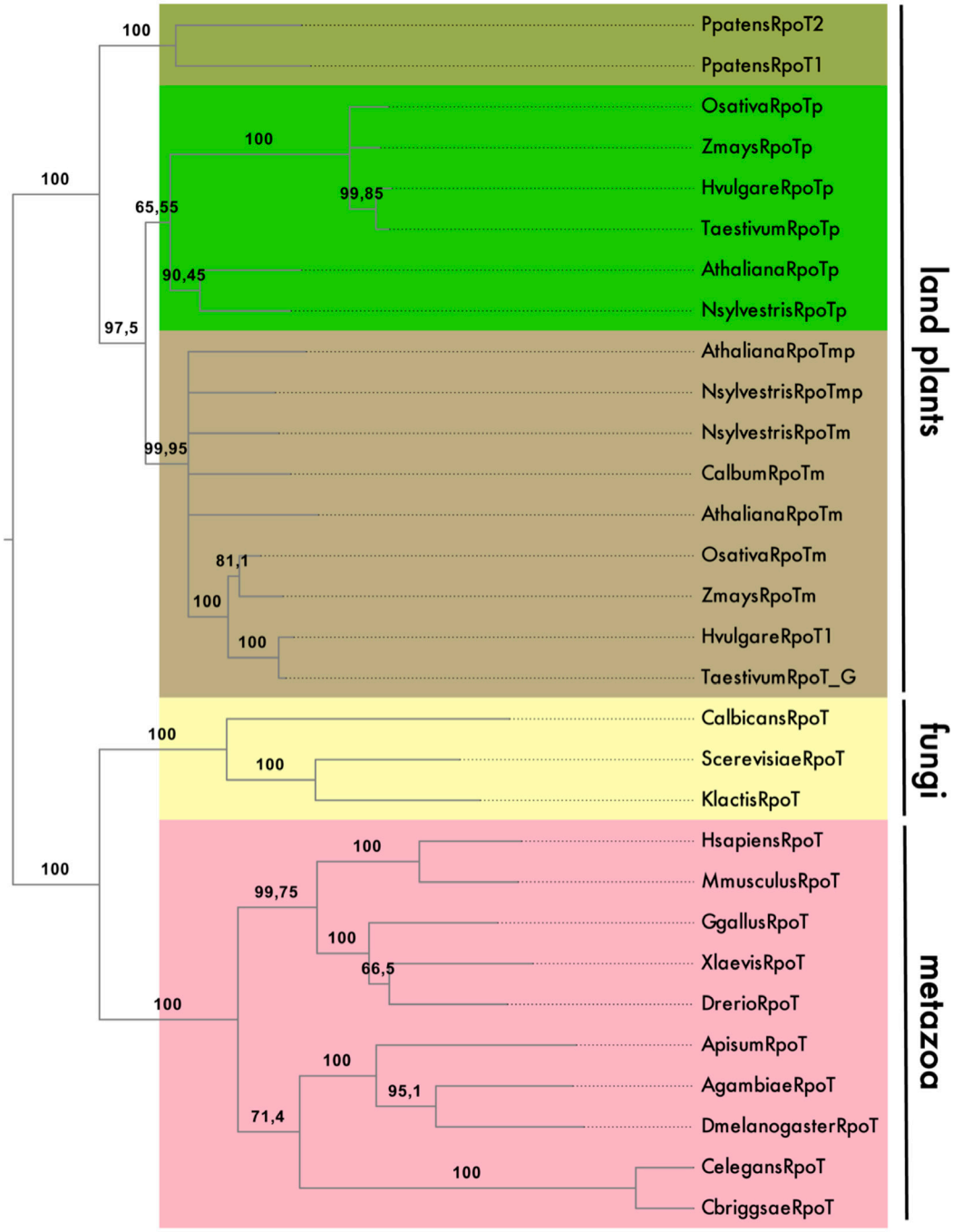


Figure 4. Schematic representation of RpoT phylogeny. *RpoT* gene duplications occurred several times during the evolution of land plants. Plant RpoT proteins form a distinct group clearly separated from RpoTs of fungi and animals. Numbers at nodes of the NJ tree show branch support values. Plastid RpoTp proteins are highlighted in green; plant mitochondrial and dual enzymes are brown; *Physcomitrella* RpoT protein clade, representing an independent duplication event, in olive; metazoan and fungal phage-type polymerases are shown in pink and yellow, respectively.

The presence of a phage-type NEP in algae has not been demonstrated thus far. Circumstantial evidence exists for two different RNA polymerases in chloroplasts of *Euglena gracilis* (Greenberg *et al.* 1984). However, an *RpoT* gene has not been identified yet nor has the enzyme been characterised (Gray and Lang 1998). Attempts to obtain homoplasmic cell lines of *Chlamydomonas reinhardtii*, lacking plastid *rpo* gene(s) have failed (Fischer *et al.* 1996), whereas such mutants exist in tobacco (Allison *et al.* 1996, Serino and Maliga 1998).

Evidence for the absence of a plastid-localised RpoT polymerase exists for *Plasmodium falciparum*, a parasite harboring both mitochondria and plastids, as analysis of the deduced transit peptide indicates mitochondrial rather than plastid targeting (Li *et al.* 2001). Accordingly, the occurrence of a plastid phage-type RNA polymerase and therefore of two types of RNA polymerases in chloroplasts (NEP and PEP) could be regarded as a typical feature of embryophytes.

1.4 The new model organism *Physcomitrella patens*

The moss *Physcomitrella patens* (Hedw.) Bruch & Schimp a widely spread moss species which colonizes open habitats in cold temperate zones has been established lately as an important model species of deep branching land plants, starting to fill the gap of plant model species between *Chlamydomonas* and angiosperms. Molecular biology and physiology concerning *Physcomitrella* has made substantial progress in recent years, starting with the major breakthrough achieved with the demonstration, that homologous recombination occurs in its nuclear DNA at frequencies which are orders of magnitude greater than those known from higher plants (Schaefer and Zryd 1997). Moreover, the gene targeting efficiency in P. *patens* is comparable with that observed in *Saccharomyces cerevisiae* (Schaefer 2002). *Physcomitrella* is therefore one of the few land plants with an efficient system for reverse genetics. In turn, the introduction of precisely predetermined DNA sequences or sequence alterations into the nucleus of the cultured organism is possible and allows studies of correlated phenotypic effects.

Several other aspects of the *Physcomitrella* biology, e.g., small size, short life cycle, a relatively easy cultivation procedure, its basal phylogenetic position, an extensive EST database and the fact that it is predominantly haploid (see Fig.4) make *Physcomitrella* ideal for molecular and comparative studies. Moreover, the whole genome sequencing project was initiated in 2004 by the *Joint Genome Initiative* (JGI) and the first assembly was published in 2008 (Rensing *et al.* 2008).

1.5 Aims of this study

The relatively low coding capacity of plant mitochondrial and plastid genomes is in stark contrast to the highly complex transcription machinery found in these organelles. While eudicots such as *Arabidopsis* use the plastome encoded bacterial-type RNA polymerase together with at least two nuclear encoded RNA polymerases of the phage-type to transcribe the plastid genome, multiple phage-type RNA polymerases are also present in mitochondria. In the moss *Physcomitrella patens* at least two phage-type RNA polymerases, PpRpoTmp1 and PpRpoTmp2 were known to be localised into mitochondria. The nuclear genes *PpRpoTmp1* and *PpRpoTmp2* encode transit peptides with mitochondrial and plastid targeting properties suggesting dual localisation of the two corresponding enzymes. The observation of multiple RNAPs in plant organelles raised the question of how these different transcriptional activities are facilitated and ultimately whether complexity of organellar transcription is a function of an increased demand for the regulation of mitochondrial and/or plastid gene expression. To study the evolution of the organellar transcriptional machineries in land plant evolution this study aimed to provide information on how organellar transcription machineries are facilitated not only in common model plant systems like *Arabidopsis thaliana* but also in deep branching land plants such as *Physcomitrella patens*.

The functioning of phage-type RNAPs in plastids of *Arabidopsis* was addressed in our group by Monika Swiatecka-Hagenbruch. To elucidate how the different transcriptional activities are facilitated in both organelles of *Physcomitrella patens* and in mitochondria of *Arabidopsis thaliana*, this study aimed to analyse mutants of the dedicated phage-type RNA polymerases.

Because studies of organellar transcription are dependent on the knowledge of mitochondrial and plastid promoters, such sequence motifs will be experimentally identified in *Arabidopsis thaliana* and *Physcomitrella patens*. Promoter mapping in mitochondrial and plastidial genomes of *Physcomitrella* will for the first time provide data on the structure of such promoter elements in a deep branching land plant species. Moreover, the experimentally determined promoters will also enable the generation of templates to test the ability of the *Physcomitrella* phage-type RNAPs to faithfully initiate transcription *in vitro*. Together with the phylogenetic analysis of plant RpoT proteins, analysis of *RpoT* insertion mutants will provide a broader view on the plant organellar transcription machineries and their evolution.

2 MATERIAL AND METHODS

2.1 Growth conditions for organisms employed in this study

2.1.1 Growth of *Arabidopsis thaliana*

For *AtrpoTmp* promoter mapping and transcript analysis *Arabidopsis thaliana* (ecotype Col-0) was grown as described in (Kühn *et al.* 2009). Otherwise *Arabidopsis thaliana* (ecotype Col-0) was grown as described in (Richter *et al.* 2009).

2.1.2 Strains and culturing of *Escherichia coli*

E. coli Top10 (Invitrogen) cells were used to propagate recombinant plasmids; Top10 cells were grown in LB or on LB agar under standard conditions (Sambrook *et al.* 2001). For recombinant protein expression growth conditions are described in 2.5.2.

2.1.3 Growth and transformation of *Physcomitrella patens*

Physcomitrella patens (Hedw.) B.S.G. was grown axenically under standard conditions (25°C; light provided by Philips TL- D 36W/25; light flux of 50 µE, light/dark regime of 16:8 h) in liquid Knop medium (per litre: 250 mg KH_2PO_4, 250 mg $MgSO_4 \cdot 7H_2O$, 250 mg KCl, 1000 mg $Ca(NO_3)_2 \cdot 4H_2O$, 12,5 mg $FeSO_4 \cdot 7H_2O$, pH 5.8) agitated on a rotary shaker at 120 rpm. Wild-type protonemata were subcultured in 7-, PprpoTmp plant material in 10-day intervals. For protoplast isolation moss material grown under standard growth conditions in Erlenmeyer flasks containing Knop medium with reduced pH (4.5) was used. Protonema was harvested by filtration through a 100 µm sieve. Cell walls were digested using 2% Driselase enzyme mixture (Serva) in 0,5 M mannitol, followed by successive passages through sieves with mesh sizes of 100 and 45 µm. Subsequently, protoplasts were washed in 0,5 M mannitol and used for the transformation procedure.

For regeneration transformed protoplasts were cultivated in Knop medium supplemented with 3% mannitol and 5% glucose in the dark for 12-16 h followed by a growth period of 10 days in the same medium but under standard growth conditions. For selection protoplasts were plated onto Petri dishes containing solidified Knop medium supplemented with geneticin (G418, 25 µg/ml), for two weeks, followed by two weeks with no selection in the medium and a second selection period of two weeks.

2.2 Generation and verification of *Physcomitrella* knock-out plants

The transformation of *Physcomitrella patens* employed the protocol described by Strepp and colleagues (Strepp *et al.* 1998). Transformation and pre-screening of transgenic plants was executed in the group of R. Reski (University of Freiburg) and supervised by E. Decker (University of Freiburg). To create knock-out constructs, fragments of *PpRpoTmp1* and *PpRpoTmp2* were amplified from genomic DNA by PCR with primers koPp1-fw/koPp1-rev and koPp2-fw/koPp2-rev, respectively (see Fig. 5) and inserted into pCRII (Invitrogen). For

the subsequent insertion of the *nptII*-selection cassette into the cloned *PpRpoTmp1* and *PpRpoTmp2* genomic fragments, plasmids were digested with BglII (pCRII-*PpRpoTmp1*) and BclI (pCRII-*PpRpoTmp2*). For transformation linearised DNA containing the *nptII*-selection cassette flanked by the genomic sequences of *PpRpoTmp1* and *PpRpoTmp2* was used. DNA and RNA samples of pre-screened Physcomitrella *RpRpoTmp2* knock-out clones were reanalysed for *nptII*-selection cassette insertion and transcript disruption, respectively. Primers employed for the generation of targeting constructs and for the characterisation of transgenic plants are shown in Table 4.

Table 4: Oligonucleotides for generation and verification of *Physcomitrella* knock-out plants

Oligo Name	Oligo Sequence (5‘ – 3‘)
koPp1-fw	gaacatatactcGAGCAGAAGGATTACGAAGAG
koPp1-rev	gaacatatactcCAATTTCAAGAAGACGGTCTAC
koPp2-fw	gaacatatactcTCGAAGTTTCCTGGCCTGATATAA
koPp2-rev	gaacatatactcGCGAAGTTACCAAACCATCTGAATTCATAAC
rekoPp2+PpiI	gaacatatactcTTCCTGGCCTGATA
rekoPp2-PpiI	gaacatatactcTACCAAACCATCTGA
testkoPp2+	CTTGGCAGCGTCAATCGAATC
testkoPp2-	CCGGTCTCTCCTAGCACGTTAAGAGC
cDNAtestko-fw	ACCGCGAGCATAGAGAGCACAGG
cDNAtestko-rev	CGCTTGGCCTGAGAAACTGCCACC
controlPp1+	AGCGAAGGAACGTGCTGCGG
controlPp1-	ATGCACGCGGCTTCACCAGC

lower case letters indicate restriction enzyme overhangs

2.3 Nucleic acids

2.3.1 Isolation of nucleic acids

2.3.1.1 Isolation of genomic DNA from *Arabidopsis*

Genomic DNA was extracted from *Arabidopsis* and *Physcomitrella* using the CTAB method (Murray and Thompson 1980).

2.3.1.2 Plasmid isolation from *E.coli*

Plasmid preparations were done following the alkaline lysis plasmid miniprep protocol (Sambrook *et al.* 2001). For large scale plasmid DNA preparations from 100-ml cultures the QIAGEN Plasmid Midi Kit (QIAGEN) was used according to the manufacturers instructions.

2.3.1.3 Isolation of total and polyA-enriched RNA

Total RNA was extracted from *Arabidopsis* and *Physcomitrella* using TRIZOL (Invitrogen) according to the protocol provided by the manufacturer. The resulting RNA pellet was washed with 70% (v/v) ethanol, resuspended in ultrapure water and DNase- treated, followed by another Trizol extraction. mRNA was enriched from total leaf RNA using the Poly(A)Purist Kit (Ambion).

2.3.2 Determination of nucleic acid concentrations

Spectrophotometric DNA quantification employed the Nanodrop ND-1000 system from peqLab.

2.3.3 Nucleid acid electrophoreses

2.3.3.1 Agarosegelelectrophoresis of DNA

DNA samples were separated on agarose gels containing 0.8–1.5% (w/v) agarose (Biozym) and 0.2 µg/ml ethidium bromide in 1x TAE buffer. For high resolution separation of 5'-RACE products, Nusieve agarose gels were prepared (1% agarose and 2% Nusieve agarose (Biozym)). Electrophoreses in 1x TAE running buffer were carried out in a horizontal electrophoresis chamber (peqlab Biotechnologie GmbH) at 7-10 V/cm. In gel visualisation of DNA molecules by UV transillumination employed the Gel Doc XR System (Bio-Rad). After preparative agarose gel electrophoreses DNA molecules were excised and purified over QIAquick spin columns (QIAGEN). 100bp or 1 kb DNA ladders (Invitrogen) were used as molecular weight markers.

1x TAE: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA

DNA loading buffer: 50% (v/v) glycerol, 1 mM EDTA, 0,005% (w/v) bromphenol blue, 0,005% (w/v) xylene cyanol

2.3.3.2 Agarose gel electrophoresis of RNA

1,2 % (w/v) agarose gels for RNA analysis were prepared with 1x MEN, containing formaldehyde (1/40). RNA samples in 1,6 volumes of RNA loading buffer were incubated at 65°C for 15 min prior to loading. Electrophoresis was carried out in 1x MEN running buffer at max. 48 Volts (5 V/cm). Separated RNA molecules were visualised as described above for DNA.

1x MEN: 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA

RNA loading buffer: 500 µl formamide, 175 µl formaldehyde, 200 µl glycerol, 100 µl 10x MEN, 2,5 µl 0,5 M EDTA, pH 8.0, 5 µl ethidium bromide (10 mg/ml), 2 mg bromphenol blue, 2 mg xylene cyanol and ddH_2O ad 1 ml

2.3.3.3 Denaturing polyacrylamide gel eletrophoresis (PAGE) of RNA

In vitro- synthesised RNAs were resolved on 0,4-mm, 5% acrylamide sequencing gels by denaturing PAGE using a Model S2 Sequencing Gel Electrophoresis Apparatus (Biometra GmbH) alongside an RNA length standard. According to the manufacturer's instructions the RNA length standard was radiolabelled before using the RNA Century Marker Template Plus (Ambion) and MAXIscript kit (Ambion). Gels were run in 0,6x TBE electrophoresis buffer, at at 55 W. After a 10-min pre-run, RNA samples denatured at 95°C for 5 min in formamide buffer were loaded and run for 2 h. Gels were dried on Whatman 3MM paper using a Bio-Rad 583 Gel Dryer. For autoradiography a phosphoimager (Molecular Imager FX, Bio-Rad) was employed.

1x TBE: 90 mM Tris; 90 mM boric acid; 2,5 mM EDTA

Acrylamide stock solution: Gel 40 (Roth)

Gel composition: 7 M urea and 5% acrylamide in 1x TBE

Formamid loading buffer (2x): 98% (v/v) formamide, 2 mM EDTA, 0,02% (w/v) bromphenol blue, 0,02% (w/v) xylene cyanol

2.3.4 cDNA synthesis and RT-PCR

cDNA was made from 1 µg mRNA-enriched or 2 µg *Physcomitrella* RNA. For cDNA syntheses the *Omniscript RT kit* (Qiagen) was employed according to the manufacturer's instructions. For nuclear encoded transcripts a mix of polydT and random hexamer primers 250 nmol each was used to prime first strand cDNA synthesis. Reactions were allowed to proceed at 42°C for 2 h. For organellar transcripts random hexamer primers were used. When PCR was employed to amplify protein expression constructs, sequences were amplified using *Phusion* DNA polymerase (Finnzyme). For macro-array analysis cDNA synthesis employed the *LabelStar Array Kit* (Qiagen) according to the manufacturer's instructions. The *QuantiTect Reverse Transcription Kit* (Qiagen) was used according to the manufacturer's protocol allowing for the elimination of genomic DNA contamination from RNA samples before reversely transcribing sample RNA. To determine the full length 5' cDNA of *PpRpoT3* 5'- RACE reactions were carried out employing the *CapFishing Full-length cDNA Premix* (Seegene) together with *Phusion* DNA polymerase (Finnzyme) according to the protocols of the manufacturers.

2.3.5 Standard PCR

For PCR amplification of genomic DNA *Taq* DNA polymerase (Qiagen) was used.

For colony PCR analysis of cloned DNA fragments, 25 µl reactions contained 0,5 U *Taq* DNA polymerase (Qiagen), 5 pmol primer each and 5 µmol of each dNTP. PCR Products were analysed by agarose gel electrophoresis.

2.3.6 Cloning and sequencing

Restriction endonucleases, Shrimp Alkaline phosphatase and T4 DNA Ligase were obtained from Fermentas and employed according to the manufacturer's instructions. When PCR products from the transcriptional start site mapping were cloned the QIAGEN PCR Cloning Kit was employed for direct ligation into the pDrive vector (Qiagen).

2.3.6.1 Transformation of *E.coli*

Chemically competent *E. coli* TOP10 cells were used to introduce plasmid DNA. Transformants were selected on solid LB medium containing the appropriate antibiotics. For protein expression the *E. coli* strain BL21 Codon Plus RIL (Stratagene) was employed.

2.3.6.2 Sequencing

The ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) was employed for sequencing according to the manufacturer's instructions. Reactions contained 250-500 ng plasmid DNA or 25-50 ng PCR fragment and 5 pmol primer. Cycle sequencing was done on an ABI 377 automatic DNA Sequencer (Applied Biosystems) and carried out by M. Meixner (SMB, Berlin). For 5'-RACE analysis of *AtrpoTmp* mutant plants sequencing of cloned 5'-RACE products was performed by Macrogen.

2.3.7 5'-RACE analysis of organellar RNAs

RNA was TAP treated and 5'-RACE was employed as described previously in (Kühn *et al.* 2005). Briefly, five micrograms of whole RNA were treated with 10 units of TAP (Epicentre), reprecipitated, and thereafter ligated to 10 pmol 5'-adapter RNA oligonucleotide using T4 RNA ligase (Epicentre). Control reactions were set up without pyrophosphatase. The adapter RNA oligonucleotide (HESWA, 5'-GAUAUGCGCGAAUUCCUGUAGAACGAACACUAGAAGAAA-3') was used for the *Arabidopsis* 5'-RACE analysis of organellar RNAs and initially also for TSS mapping in *Physcomitrella patens* as described before (Kühn *et al.* 2005) . After the initially used adapter RNA was finished up, a new RNA oligonucleotide was employed (5'-GUGAUCCAACCGACGCGACAAGCUAAUGCAAGANNN-3') which allowed for the nesting of two forward primers (RUMSH1 and 2). Additionally, the last three nucleotides of the RUMSH RNA adapter were set to N. The latter feature was introduced to allow for a statistically more significant mapping of the 5' TSS, since it would allow to identify multiple independent RNA ligation events. In all cases transcripts were reverse transcribed by Superscript III (Invitrogen) or Omniscript (Qiagen). Of note, gene-specific primers were used for reverse transcription (RT). Up to 5 gene-specific primers were used in one RT reaction. Products of reverse transcription were amplified in a first PCR step by using 1 to 3 µl of the RT reaction. Adapter primer P1a (see Tab. 5) served as forward primer in PCR reactions for the HESWA RNA oligonucleotide ligation products and adapter primers RUMSH1 and RUMSH2 (see Tab. 5) for the RUMSH adapter ligated products. Gene-specific reverse primers are provided in Table 5. Products ligated into pDrive (Qiagen) and transformed into *E. coli* TOP10 (Invitrogen). Bacterial clones were subjected to colony PCR using primers

M13-seq-F and M13-seq-R. The amplified products were purified and sequenced.

Table 5: Oligonucleotides used for mapping and sequencing of organellar 5' transcript ends

Oligo Name	Oligo Sequence (5‘ – 3‘)
M13-seq_fw	ACGACGTTGTAAAACGACGG
M13-seq_rev	TTCACACAGGAAACAGCTA TGAC
P1a	CGAATTCCTGTAGAACGAACACTAGAAG
RUMSH1	TGATCCAACCGACGCGAC
RUMSH2	ACCGACGCGACAAGCTAATGC
AtP2-cox1	TGTGCCCATCACTCCAG
AtP3-cox1	ACCGAAAATGAAATAGAGAGTCCCT
AtP4-cox1	TGTGGTTTGTGGAGAACAGCC
AtP4-cox1-b	GAGTTCTTGAGAGTCCCGTGG
AtP5-cox1-b	ATCGTCCTACAAAAGATAATGCTCTCAC
AtP2-rps4	TTGAAATCTTAATGGGGACA
AtP3-rps4	CAGCCACATTGCGTCCC
AtP4-rps4	GGACTCGAACCCAATTCTGC
AtP5-rps4	CATGACTCGACCTTTAGTTTTGTGAC
AtP6-rps4	GAATTCGCAAGACACCATAGGC
AtP7-rps4	GAAAAGGTCGGGACAGTGGC
AtP2-rps4-b	GTTTTGTTCCTCTGTGCATC
AtP3-rps4-b	TGCGGCGTTGTATTATTGTCAGT
AtP4-rps4-b	CAAGTTTGAAATCTTAATGGGGACA
AtP4-cox1N	GATTCCAGTGGGGACAGCTATG
AtP3-cox1N	ATTCCAGTGAGTCCTCCTATGGTG
AtP2-cox1N	GCATACCTGAAAGCCCTAAG
AtP5-nad1e5	GTCCAAGCGATAGCGTAAAAGC
AtP4-nad1e5	TCTAGGATAGGCGGCCAACC
AtP3-nad1e5	TTTCCGGCCAAGTCCCA
AtP2-nad1e5	AAAAGACCAGAAACAGGGAC
AtP6-nad1e5	GGGCCTTTTCTTGATCGTTAGC
AtP7-nad1e5	TCAGGTCTTGACCGGGTCC
AtP4-matR	TGGAAACTCGGGATCGTAAATG
AtP3-matR	CGGAAATGCGATGTGTCTGG
AtP2-matR-a	GAGTGGCAGCCTTGACC
AtP2-matR-b	TGCTTGTGGGCTAGGGT
AtP3-matR-b	CAATGACAACTTACGGGCATGAG
AtP4-matR-b	CTTGCCGGATGTGCTTGA
AtP5-matR-b	GCTAGTTGAACCCCGTCCTGT
AtP3-nad1e4	CGCACCTCATTAAGATCATATTGG
AtP4-nad1e4	AAAAGAGCAGACCCCATTGAAG
AtP5-nad1e4	CTCTCGATGTGAGATCAGCAGC
AtP6-nad1e4	CCCTTGCCCTTGAACGATAGA
AtP4-nad1e4-b	GATCATTACGAGCCCTTTCCTAGA
AtP7-nad1e4	ATCTCGTAGCACCACAGCCTG
AtP5-nad2e1	GAAAAGAAAGGGCGGAATAGCA
AtP4-nad2e1	GTGGGAGGATCCGAACGA
AtP3-nad2e1	TCCAAGCCAACCCACATTACTG
AtP2-nad2e1	AAGTAATCCAAGCCAACCC
AtP4-nad2e2	TAGGAGAGGTGCGCCAGC
AtP3-nad2e2	CATCAAACCTCTCTTGGTCGG
AtP2-nad2e2	AAAGAGCATACCGCGAGTAG
AtP6-nad6	CACGGAGCGGTAGACTGAACAC
AtP5-nad6	AAACCAACAATCTGCACGCTTAG
AtP4-nad6	GACCAAAGCAGGGCTCGAC

Oligo Name	Oligo Sequence (5‘ – 3‘)
M13-seq_fw	ACGACGTTGTAAAACGACGG
M13-seq_rev	TTCACACAGGAAACAGCTA TGAC
P1a	CGAATTCCTGTAGAACGAACACTAGAAG
AtP3-nad6	GATGGGAAACAAAACGGAATGT
AtP2-nad6	CGCAATACTTCTTCGTGAA
AtP2-ccmC-a	AGCGAAGATAAAGGGGATG
AtP3-ccmC	TGTTAAGAACAACCAAGACCCAATG
AtP4-ccmC	AATGAGAATTTGCGCGTAGCTT
AtP5-ccmC	TGCTGCAAAAGACTGGGATG
AtP6-ccmC	GGGATGGGAAACAAAACAGAATG
AtP7-ccmC	CAGAACAAAGCAGGTAGACCAAGC
AtP8-ccmC	CCGTCTTTCAATGCCTGTACC
AtP9-ccmC	GCGTCGAGTAAGAGCATCTGG
AtP2-ccmC-b	TTGTCTATAATAGGAAGCGAACA
AtP4-nad9	CGTATTGGTATCAGATCTATTCCCATGT
AtP3-nad9	TGAAATAGGTAGTCCGTATTGGTATCAG
AtP2-nad9	AAAGCACAACAATTGAAATAGGT
AtP5-nad9	GCTTTTGAGATGCCCTAGACGA
AtP6-nad9	TTTGCCAAACTCACACCATCC
AtP2-nad5e2	GAAATCCCAAGAGCTAATCC
AtP3-nad5e2	AAGACCTACTCCCTCCCATCCC
AtP4-nad5e2	CTATGCGGATCCTCGGACA
AtP5-nad5e1	CGAAGAGAATGAAACGCACGTAG
AtP2-nad4-a	AGAGAGGCACACAGACCAA
AtP2-nad4-b	TGTACTGGCTTCCGCACTA
AtP3-nad4	CACACAGACCAATTAATCGTATCGG
AtP4-nad4	GCACAGGACACAGAATAAGACCAC
AtP6-nad4	CAAAATGTGAGGGAAGCACTAAGC
AtP5-nad4	GTCGGATCAATCTCACTCTCGAA
AtP7-nad4	CACCGGAGCTATCTGCGTC
AtP8-nad4	CTTAGCGGCACCTCTGACCT
AtP2-orfX	CAGCCCTCTTCACGAACT
AtP3-orfX	CGACGATTTGGCACCAGAT
AtP4-orfX	TGTGGAAAGAGCAGCTGTGAGA
AtP5-orfX	TTCCGGGAACCAGTAACACGTA
AtP4-ccmFc	CCATTTCAATAGTACGGGTGCTG
AtP3-ccmFc	TCCACTTTTTGCTCCGTCCA
AtP5-ccmFc	ACGAGTAGGTAGGCGGCATCT
AtP6-ccmFc	TAGGAATCTTAGGTTACCTCCAGCC
AtP2-ccmFc	TGAAGAACGGGAACGAA
AtP3-rpsl2	AACACGCGGGCATACTCCT
AtP4-rpsl2	TCCAAAGCTCGAGTACGGTCC
AtP5-rpsl2	CAAATGGAAAAGGAACACCGAGT
AtP7-rpls2	GGGCAACAACAGTCGAACC
AtP6-rpsl2	CAAAAACTACCTGCCAAAGCTCC
AtP8-rpsl2	GAAGAACGGCCTGCGAACT
AtP2-rpsl2-a	GTGGAGCGGAATTAGGTT
AtP2-rpsl2-b	CGCTTCCTTCTGATCCTAGA
AtP2-rps3	GCCCTCACTGAACCGACT
AtP3-rps3	GAACCGACTTGAATCTGAACTACGA
AtP4-rps3	TTTTTGACTTTATGGATTTCTGTCCCT
AtP7-rps3	AGATAGAAATGATAGAGGGCCAACC
AtP2-atp6-1	GGGATCTTGCGTTAATGC
AtP3-atp6-1	GATCTTGCGTTAATGCCTCACAC
AtP4-atp6-1b	CAAACAAAAAGATTCGTCGCATATTG
AtP7-atp6-1	CGGTTCATCGCCTTACTTATCCA
AtP2cox2-c	CCGAAGAATCTCGATAGTAG
AtP3cox2	GTAGCTGCGTCTTGAGATCCTAATTG

Oligo Name	Oligo Sequence (5‘ – 3‘)
M13-seq_fw	ACGACGTTGTAAAACGACGG
M13-seq_rev	TTCACACAGGAAACAGCTA TGAC
P1a	CGAATTCCTGTAGAACGAACACTAGAAG
AtP3cox2-d	TTCTTCTTCTTCTTACAATATTTTGAGTTAGATG
AtP5cox2-d	CGAAACCAACATCCTTATAATACTACTAGGC
AtP7cox2	CATTAGATAGCTAATTATCCTTTGCCTAGC
PpaccDrev3	CAATTAATCCACCGAATCGCC
PpaccDrev2	CCTTTCTCGTTCATTAAGAATGTAGC
PpaccDrev1	TCACAACTATCACATCGAGTCC
PptrnGrev1	GAATGATTAAAAGGCGGGTAGC
PptrnGrev2	AATCGAACCCGCATATTCTCC
PptrnGrev3	TTGGCAAGGAGGAATTTTACC
PppsaArev1	CCATGAAAATACATACCACTTAGC
PppsaArev2	CCAAGTTGTAGTATTAGGACCC
PppsaArev3	AAACCCACCAATAATGAATAGTAAAC
PppsbArev1	CAGTGCTGGTAACCCAGTCG
PppsbArev2	CATAGGCTTGCGCTTTCG
PppsbArev3	GTAGCAGTCATGGTAAATCTTGG
Ppycf2rev1	TCCACCATTCCGAATCCCCAAC
Ppycf2rev2	AATTTGATTCTGACCATACTTCAAA
Ppycf2rev3	ATCGAGGATTTTGTATTTCTTCCAACT
PprpoBrev1	TTTCGACGAAGCAATCGACATG
PprpoBrev2	TGAATTTGTCTTAATTCAGGAAGTG
PprpoBrev3	GATAAAAATGAAAATTGAGATGGG
PpatpBrev1	ACACCACCGTGCGCTTTAGCAA
PpatpBrev2	GTCGCACTCATTGCTACAGCTCG
PpatpBrev3	TGTTGAAGCTCCAAAAGTACGGG
Ppnad6rev1	TTGTCGGAACACATCCTGTCT
Ppnad6rev2	ACCACCTACAGGTAAATAACGCA
Ppnad6rev3	GCAATAGCTCCTACATAAACC
Ppnad6rev4	GCTGCTGGAGCAGAGGAAGGG
Ppnad6rev5	TTCGCCCGCGAAACCCGTAA
Ppnad6rev6	GCATCACCGGAGGGGAGGCT
PptrnEmtrev1	TCGAACCCGTGTTCTCGCCA
PptrnEmtrev2	GCGGCCTGTGGCCCGTTTAT
PptrnEmtrev3	ACCAAGAGCGACCTTTCGTCTGC
PptrnEmtrev4	ACCGTCGCTTAAGTGAATAGGGGA
Ppatp6rev1	CCTATATGAATAGGAATCAATGG
Ppatp6rev2	ATTGCAAATTGTTCTAGCGGAC
Ppatp6rev3	TAGCGGACTGCAAGCCATG
Ppatp1rev1	ACACCGCTGGCAAATTCAACCA
Ppatp1rev2	ACCACTCGACCGATCTCATCCACT
Ppatp1rev3	TGCTCCAGCCAACTTATTTCCAGT
Ppatp1rev4	GCCAATGAAGTAGCGGAGGCCC
Ppnad9asrev1	CGGCTGGTGGGAGCGAGAAG
Ppnad9asrev2	TGCTCGGCTCAGTTCTATGGCA
Ppnad9asrev3	CGCTCGGGTGATGGGGATTCG
Ppnad2rev1	ACCAAGCCAACTAACATTACACACT
Ppnad2rev2	TGCGTTAATAAGAAAGATCTCTGGG
Ppnad2rev3	AAGATCTCTGGGAAAAGCGC
Pprrn26rev1	CCCCATGGCGTTTCGCCCTT
Pprrn26rev2	AGCGTCCTTCCTTCTCAATGC
Pprrn26rev3	TTGGATACAGTAGGAATTGCACC
PptrnfMrev2	CATTCAGATTATGAGCCTGACG
PptrnfMrev3	CCGCGCACATAATAAGGT
PptrnfMrev4	TCTATGATGATTCATTCGTAGGC
Pprrn5rev2	TCCGGGCCTGGACCATGTCT
Pprrn5rev3	GAAGCGCGTAGCGATTAATG

Oligo Name	Oligo Sequence (5‘ – 3‘)
M13-seq_fw	ACGACGTTGTAAAACGACGG
M13-seq_rev	TTCACACAGGAAACAGCTA TGAC
P1a	CGAATTCCTGTAGAACGAACACTAGAAG
Pprrn5rev4	GAAAGCAAGCCATGTAGACC

2.3.8 Northern

2.3.8.1 Preparation of riboprobes

For the specific evaluation of organellar transcript abundances single stranded RNA probes were synthesised with Ambion MAXIscript kit (Applied Biosystems) according to the manufacturer’s instructions. To generate probe templates for organellar genes from total DNA Phusion Hot Start high-fidelity DNA polymerase (Finnzyme) was employed. Purified PCR products were used for in vitro transcription reactions (see above).

Table 6: Oligonucleotides for riboprobe synthesis

Oligo Name	Oligo Sequence (5‘ – 3‘)
AtmatR-t7-F	GTGGCCTGTTTGACGAACGTCAG
AtmatR-t7-R	taatacgactcactatagggTTGTTCAGGTCTTGACCGGGTCC
Atcox1-t7-F	GTAGGTAGCGGCACTGGGTG
Atcox1-t7-R	taatacgactcactatagggATACCGAATCCAGGCAGAATGAG
Atrps4-t7-F	CAAGCAAGGCAGCCGATAAGTC
Atrps4-t7-R	taatacgactcactatagggATTCTCTTGTGCTCAGCGAAGGA
AtccmC-t7-F	TTGGTTGTTCTTAACAGCGATGG
AtccmC-t7-R	taatacgactcactatagggAAATAGAAGCCGGTTCGACAGG
AtccmFc-t7-F	ACTATTGAAATGGTTCGTCAGTAGAGATG
AtccmFc-t7-R	taatacgactcactatagggCGCTTCGCTGACCTATCGC
Atnad9-t7-F	TGGAAAGATCGGAACATGGGAATAG
Atnad9-t7-R	taatacgactcactatagggCTTGGGTCATCTCAATGGGTTCA
Atnad2-t7-F	TGGCGCACCTCTCCTAACTATTG
Atnad2-t7-R	taatacgactcactatagGGCTTCCGTGGAAAATTCAGACT
Atnad3-t7-F	GCTAGTTTCTTTGATCCTACTCGGTG
Atnad3-t7-R	taatacgactcactatagggACAGATCAATCTTGTTGGGAGGTAC
PpT3Frnaprob	GTGGAGTGCAATCCTTTGGT
PpT3Rrnaprob	taatacgactcactatagggGCAAGGAGTACCGCTCGTAG
PpT2Frnaprob	CTACGACCAGCATTTCAGCA
PpT2Rrnaprob	taatacgactcactatagggTACGCTTGGCCTGAGAAACT
PpT1Frnaprob	CCTGGGGACCTGCTATACAA
PpT1Rrnaprob	taatacgactcactatagggTTTGCTGAGGCAGTATGTCG
Ppcox2Nprobe+	GGACTGGCCGTCGACCCAACC
Ppcox2Nprobe-	taatacgactcactatagggTCGGGCGCACGCATAAACGC
PpccmF-Nprobe+	AGCGCCAAGCGCTTTGTTGGC
PpccmF-Nprobe-	taatacgactcactatagggCGCATAAGCTGCCGGGTCG
PpccmFN-Nprobe+	GCCAATACGAAGGGAAAGCAAGGCCG

PpccmFN-Nprobe-	taatacgactcactatagggTTCGCGCTGCGGGTCCAACC
Pprps1Nprobe+	CGGTCAAAGGAGGTTATGCGGTAGC
Pprps1Nprobe-	taatacgactcactatagggTGTGGGCTTGAGGCTTTGGGC
Pprpl5Nprobe+	CACGAAGCAGAGATTCGGCAGG
Pprpl5Nprobe-	taatacgactcactatagggACGACGTTGGCATTGATAATTGAATGG
PpatpBNprobe+	ATGAATGAGCCGCCTGGAGC
PpatpBNprobe-	taatacgactcactatagggCTTTCGCTGCTAGCCCTCTGG
PprpoBNprobe+	CGTCGTATCCGCTCTGTAGCCG
PprpoBNprobe-	taatacgactcactatagggTCTTGTCGATAGCGAGCTGGAGT
Pprrn18Nprobe+	AGCAGCCGCGGTAAGACGGG
Pprrn18Nprobe-	taatacgactcactatagggTCCACCGCTTGTGCAGGCCC

lower case letters indicate T7 promoter sequences

2.3.8.2 Detection of RNA

Total RNA was resolved on 1,2% agarose-formaldehyde gels and transferred onto Hybond N+ nylon membranes (GE Healthcare). For hybridisations with in vitro-biotinylated complementary RNA probes (*Arabidopsis* northern) the method described by Falcon de Longevialle et al. was employed (de Longevialle *et al.* 2008). The chemiluminescent nucleic acid detection kit (Thermo Scientific) was used for probe detection on an ImageQuant-RT ECL system (GE Healthcare).

Radiolabeled ribonucleotides were used for RNA probe labelling and detection of organellar transcripts from *Physcomitrella.* Hybridization was performed for 16-20 hours in UltraHyb buffer (Ambion) at 48°C. Membranes were washed twice in 2x SSC supplemented with 0,1 % SDS for 20 minutes and additionally 1x in 0,2x SSC/0,1% SDS for 15 minutes. Dried membranes were exposed 16 to 72 hours on Imaging Screen K-Type (BioRad). Visualisation employed the phosphoimager (Molecular Imager FX, Bio-Rad) and the QuantityOne software package (BioRad).

2.3.9 real-time-PCR

Primer pairs for quantitative real-time PCR of cDNA samples were designed to give amplification products from 70 to 120 bp. The qRT-PCR was carried out in a 7500 Real- Time PCR System (Applied Biosystems). Using the TaqMan Fast Universal PCR Master Mix (Applied Biosystems) or the qPCR MasterMix Plus (Eurogentec), each reaction contained 50 ng of cDNA, 100 nm of the particular probe and 1 µM primer each.

The cycling was performed as follows: 95°C for 10 min, 40 cycles of 15 s at 95°C and 1 min at 59°C. Removal of genomic DNA from every cDNA product was verified consistently through a -RT control reaction (no reverse transcriptase added). All samples were analysed in triplicates and none template controls were included for each primer pair. Data analysis employed the Sequence Detection Software v1.4 from Applied Biosystems. Normalisation of acquired signals were normalised to the amount of 18S transcript as an internal standard unless stated otherwise using the ΔCT method (2(-ΔCT) = relative amount of transcripts; ΔCT = CT target – CT internal standard). Used primers are included in Table 7.

Table 7: Oligonucleotides employed for qRT-PCR

Oligo Name	Oligo Sequence (5‘ – 3‘)
rrn18nc+probe137	TGTGGGTGTGCACTGGTC
rrn18nc-probe137	AAAGTAACATCGCCGACTCC
atp6ex2+probe165	ACTCCTTGCGGCAATAAAAA
atp6ex2-probe165	TCATTATCTCTCTTTATCGGAATCAC
cox1ex1+probe31	CACAGCCTGGCAATCAAAT
cox1ex1-probe31	GCATAACCATGAAAAAGATCATTAAA
cox1ex5+probe125	GTTTCGGCTCGTATGTTTCTG
cox1ex5-probe125	CAAGTGTCGTTGAATTCTGTTCA
matorf533+probe94	GCTAGCCAACATTGCTCTCC
matorf533-probe94	TGGGATATGCTCTCAGTGACC
matorf622+probe15	TGGATAACACTCCTTGCGTCT
matorf622-probe15	TCACGAGATCCTCGGAACAT
nad2ex2+probe7	AACCCATGGATCGTGAAAAA
nad2ex2-probe7	TGCCATTTGATGGCTGACTA
nad6+probe67	CTTATTTTGCACCCCCTTCC
nad6-probe67	GCATAAAACAAGAGCCCCATA
nad9ex1+probe12	AAAGCGCTGGCACTGTATG
nad9ex1-probe12	GGCTAGCAGGGGATACGTC
rps12+probe72	CCAATCGAAATGAGATAATTGCT
rps12-probe72	TCGCTCTACCCCCTCTAACC
trnaGlu+probe106	CCGCAAATAGTGCCACTTC
trnaGlu-probe106	CCCTAGACGAAAGGGACAAA
PpRpoTmp1+probe101	ATTTCGGCGGTTGGTAGG
PpRpoTmp1-probe101	GCGCAATCCAACCAGTTTA
PpRpoTmp2+probe39	AGTCGGGATTGCGATTCTT
PpRpoTmp2-probe39	TGCGGAATCTAATGGTGTCA
PpRpoTmp3+probe103	TGCATTCTATTTGCCACACAA
PpRpoTmp3-probe103	GGTTAAGATTTGGGTGCATAGG

2.3.10 Preparation of Macroarray membranes

Whole RNA from *Physcomitrella patens* was reverse transcribed and cDNA fragments of organellar and nuclear transcripts were PCR amplified, cloned into pDRIVE (Qiagen) and sequenced. Subsequently, PCR from verified clones was performed. After gel purification of amplified fragments 2 µg of each of the PCR products was dissolved in water.

PCR products, 1 M NaOH and 5 M NaCl were mixed in a ratio of 1:1:1 and spotted onto nylon membranes (Biodyne B, Pall, Dreieich, Germany). Each PCR fragment was spotted in duplicates. BioGrid robot (Biorobotics, Cambridge, UK) spotting procedure employed five transfers to each spot with solid pins of 0,2 mm diameter. After spotting, arrays were washed with 0,4 M NaOH, 1,5 M NaCl and neutralised with 0,5 M Tris-HCl, pH 7,5, 1,5 M NaCl.

Spotted DNAs were fixed on the membranes by UV-Crosslinking at 120 mJ in a Stratalinker (Stratagene). Subsequently, membranes were washed with 2x SSC and dried for 30 min at 80°C. Arrays were stored at room temperature. For technical replicates filters were stripped by shaking in boiled 0,1x SSC, 0,1 % (w/v) SDS for 10 min at RT and subsequent denaturation in 0,4 M NaOH, 0,1 % (w/v) SDS followed by neutralisation in 0,2 M Tris HCl, pH 7.4 (2 x 15 min) and 15 min washing in 0,1x SSC, 0,1 % (w/v) SDS. After stripping cross-wise hybridisations of wild-type and PprpoTmp cDNAs were performed on these filters.

The signal intensities were determined using the image processing software ArrayVision 5.1 (Imaging Research Inc.).

Table 8: Oligonucleotides preparation of Macroarray membranes

Oligo Name	Oligo Sequence (5‘ – 3‘)
Pp-coxIIIfw	GGATGATTATGTTTCTTGGGTA
Pp-coxIIIrev	AATTCGGAAGGATTTAACACA
Pp-nad2fw	ATCTCGGCTTATGATTTAATTG
Pp-nad2rev	GGCGAATACATTGATGGT
Pp-nad5fw	ATTCACGCAGCTACTATGGTAA
Pp-nad5rev	CCAACCAGAAAGCAAAGTTA
Pp-nad7fw	CTGTGGGCCTTTGAAGAG
Pp-nad7rev	TTAGGTGCTTCTACTGCGGTAT
MP-rrn26fw	ATAGGTGGGAGGTGGCGACA
MP-rrn26rev	TAGGGTCGGCTTCTCGCAGTT
MP-atp9fw	AATTGGAGCAGGAGCAGCTA
MP-atp9rev	TGCCATCATTAAGGCAAACA
MP-atp1fw	AAAGGGCATGTTAGGTCGTG
MP-atp1rev	CGCCACTGATTGCTTACTCA
MP-cobfw	TATACCCGCAAATCCCATGT
MP-cobrev	GGGTTGACATCCAATCCAAC
MP-nad3fw	GCTTCTTCTTCCAGTTTGGC
MP-nad3rev	ATCTAAAGCGCCCTTTTTCC
MP-nad4Lfw	TTTTTCTTTTAGGTATTTGGGGAA
MP-nad4Lrev	CACTGCAATAGTACCGCGAA
MP-nad6fw	ACGTGCCAAAAATCCAGTTC
MP-nad6rev	TTGCACTCAATTTCGTTGGT
MP-nad1fw	TTTTCTAATGGCACCCGTTC
MP-nad1rev	GTAGCCCGCTACGAGTTCTG
MP-rpl6fw	TTTTTGGAAATAATTGGAGTTGG
MP-rpl6rev	CTTTTGACAATGGCAGCAAA
MP-rpl16fw	AGGTTGCAAAGCAGACGGTA
MP-rpl16rev	GCAGCTTGTTGAGCATTTGA
MP-rps2fw	GACGCTTTTACATCTTCGCC
MP-rps2rev	GCCTCTGCGATCTTTTTGAA
MP-rps11fw	TTGGTCCTCCTCAGGTTCAG
MP-rps11rev	ATTATGTGGCGTTGGGGTTA
MP-trnArgfw	GCTCAGTTGGATAGAGCAACAA
MP-trnArgrev	CGCATCCTACAGGATTTGAA
MP-trnGlufw	GTCCCTTTCGTCTAGGGGTA

MP-trnGlurev	GATTTGAACCCGTGTTTTCG
MP-trnHisfw	GGCGGATATAACTTAGGGGTT
MP-trnHisrev	ATTCGAACCCGTGTTTTCAG
MP-trnLysfw	GGTGTATAGCTCAGTTGGTAGAGC
MP-trnLysrev	GACTTGAACCTGCGACCTTT
MP-trnfMetfw	CGCGGGATAGAGTAATTGGT
MP-trnfMetrev	GGATTCGAACCCACAACATT
MP-trnSerfw	TGTCTGAGCGGTTGAAAGAG
MP-trnSerrev	GAACCCCCGGTATTCTCAAT
Pp-mtMnSOD-p	GCGGAGGGCATGTTAAT
Pp-mtMnSOD-m	TTTCTGTGGTGCGATACAAT
Pp-mtPeptase-p	GGATCAGAGGTTCGGATT
Pp-mtPeptase-m	GTCATAAATGAAGCGACTGG
Pp-mtTOM-p	CACTGACGCCGACAATTT
Pp-mtTOM-m	ACCAACTGCGAGTACAATCC
Pp-mtMdh-p	CAAGCGGACTCAAGATG
Pp-mtMdh-m	TACCCGTAAGATGACAAGAT
Pp-mtAoxcDNA-p	CTTCAACGCGTACTTTCTTC
Pp-mtAoxcDNA-m	CCCGCAGCTCCTTACTT
Pp-CycD-p	GTCAGTGGCGTGCATATCTC
Pp-CycD-m	CATGCGGACACAGTAAGGA
Pp-rpl2-p	GCGGAACTACTCGTATCAAA
Pp-rpl2-m	CCAACTAGAGCGGAGCAT
Pp-18S-p	GGGCATTCGTATTTCATTGT
Pp-18S-m	AGACGCCGATAGTCCCTC
Pp-25S-p	AGACTACCCGCTGAGTTTA
Pp-25S-m	AGCCTTAGAAGGAATTTACC
Pp-ptAsx-p	CAGTTTCCGGAGTTGTCGTA
Pp-ptAsx-m	CATGTGAGACTCGGCGTAG
Pp-ptCab-p	TTTGGGATTCGGGAAG
Pp-ptCab-m	TTTCCTAAGACCTCGTCAAC
Pp-ptAtpD-p	GCGAGCTGATGACTAACGA
Pp-ptAtpD-m	ACAATAAACCCAGCGATCAA
Pp-ptCcsA-p	AGCCAATTCAGTCGAACATC
Pp-ptCcsA-m	ACAGTTACCGCAGCATTACC
Pp-ptFtsZ1-p	TATGGTGGAAGAAGCCCTAC
Pp-ptFtsZ1-m	TTGCGTTGGGATCGAC
Pp-ptFtsZ2-p	GTGACGGAAGCATTCAATCT
Pp-ptFtsZ2-m	CCAATACGCGACTTGCATA
Pp-ptHxk-p	GATGCTTCCGACCTATGTC
Pp-ptHxk-m	TACCAGGGCTGAGATCATAA
Pp-ptLhcP-p	CATATCTGAATGGCGAGTTC
Pp-ptLhcP-m	CCAAGTGGTCGTTCAAGTT
Pp-ptOee2-p	CCCAAGAAGCAGACAGGT
Pp-ptOee2-m	CTTGCCTCTTTCATCTCCAT
Pp-PhyA-p	ATCCAGCCGCACACCT
Pp-PhyA-m	CCTCACTAACGGGAAGTCCT
Pp-PhyB-p	CAAGCAGTTACCGCATACCT
Pp-PhyB-m	CCGTGTTCGTCTTCGTG
Pp-ptPsaD-p	CGCAGGTTGAGGAGTTCTAT
Pp-ptPsaD-m	GAGCTCCGCAGGGTTT
Pp-ptPsaF-p	GCGGGAGTGATATGGTTGT

Pp-ptPsaF-m	GTCAATCAAGTAGGCCCTGC
Pp-ptRpoA-p	TCTGGTCAAGCGAACACTG
Pp-ptRpoA-m	TTCTTGCGTTACGTCAGGTT
Pp-ptSig1-p	GATAAACCTTCGCCTCCTG
Pp-ptSig1-m	TATCTTGCTTCCGCTCAACT
Pp-ptSig2-p	CCGGAGTATGTTACCTTCAA
Pp-ptSig2-m	TTCTCTGCCCTTTCCAA
Pt-ycf2_fw	ATTGGATAAACCGCCCTAATGA
Pt-ycf2_rev	TTGAATCTCGGCCATATCCAC
Pt-ycf1_rev	CCTTGGGAACCGCTAA
Pt-ycf1_rev	ACGTCTTCTAGAGCGCATAG
Pt-trnV-GUCfw	GGGATATAACTCAGCGGTAG
Pt-trnV-GUCrev	GGGATAATCAGGTTCGAACT
Pt-trnS-UCCfw	AAAGATGGCCGAGTGGTC
Pt-trnS-UCCrev	GGAAAGAGAGGGATTCGAAC
Pt-trnS-UCAfw	AGAGATGGCCGAGTGG
Pt-trnS-UCArev	GAGAGAGGGATTCGAACC
Pt-trnS-AGCfw	GAGAGATGGCCGAGTGG
Pt-trnS-AGCrev	GGATTCGAACCCTCGGTA
Pt-trnP-CCAfw	GGGATGTAGCGCAGTTT
Pt-trnP-CCAfw	TAGGGATGACAGGATTCG
Pt-trnN-AACfw	CAGTGGTAGAGCGGTCG
Pt-trnN-AACrev	CTCCCCAGGTAGGACTTG
Pt-trnL-UUGfw	GCCTTGGTGGTGAAATGGTA
Pt-trnL-UUGrev	GCCTTGAAGAGGATTCGAAC
Pt-trnL-CUAfw	TTGGTAGACACGCTGCTCTT
Pt-trnL-CUArev	CCGAGATGCGCTAGCACT
Pt-rrn23fw	GGTTAGCCGAAAGATGGTTA
Pt-rrn23rev	GAACTTACCCGACAAGGAAT
Pt-rrn16fw	CACCAACGGCGAAAGCACT
Pt-rrn16rev	CCAACACCTTACGGCACGAG
Pt-rps15fw	ATTGGATCATCTTTACTTTCAA
Pt-rps15rev	GACGCTTACGTTTTCCTAATA
Pt-rps12fw	CGCCTTACGTAAAGTAGCAAGAG
Pt-rps12rev	GAACGCCCTTGTTGACGA
Pt-rps4fw	GAATAATACGTCGTTTAGGAGTT
Pt-rps4rev	CCTCAGGTTTACAGCGATAA
Pt-rps2fw	GAAGAAATGATGCAAGCAGG
Pt-rps2rev	CATCATTTGCTGGAATTGGT
Pt-rpoC2fw	TTCGCTCACCTTTAATTTGT
Pt-rpoC2rev	ACGAACTTCCGCAATAACT
Pt-rpoC1fw	CTGCCAGTTCTTCCTCCTGA
Pt-rpoC1rev	AGCAAGATGGCGTCCAATTA
Pt-rpoBfw	AAGATGGAACGCCTATTGAT
Pt-rpoBrev	CCGACTCTTTGTCCTCCTC
Pt-psbDfw	TTTATGTTGCGCCAATTTGA
Pt-psbDrev	TGATCTTGAGCTGCCATCC
Pt-psbAfw	TTTGCACTTCTACCCAATTT
Pt-psbArev	ACAGGCCAAGCAGCTAA
Pt-psaJfw+	AACATACCTTTCTACAGCACCC
Pt-psaJrev+	AAGGAAGAACTAAAGCATCTGG
Pt-petDfw	TTATGGAGAACCTGCTTGGC

Pt-petDrev	CTACTGGACGACGGAAAGGA
Pt-petBfw	CGATTGGTTTGAAGAACGTC
Pt-petBrev	CCAATTTGATCCCAAGGTAA
Pt-petAfw	TATTTCCGAAGCATATCCAA
Pt-petArev	GGTCCTGCCGGTACAA
Pt-ndhKfw	CCGGCCTTTTCTAATCAAGA
Pt-ndhKrev	TGGTTTCGGCGGACAA
Pt-ndhHfw	CTACTGGTATGCGAATGATG
Pt-ndhHrev	AAGTCGTCGAGCTTCTAAAT
Pt-ndhFfw	GGACGCTATGGAAGGAC
Pt-ndhFrev	CGAGCCAAGAGTAGAAGAAT
Pt-ndhDfw	TATATACGGCTGGTGGTTCC
Pt-ndhDrev	GCGCCATTAAGACCTACATC
Pt-ndhCfw	TCAGGAATAGAACCAATGGG
Pt-ndhCrev	CGCCATGCATAAACTAAACC
Pt-ndhBfw	TTAACTGCAACTATCGGAGGAA
Pt-ndhBrev	AAGTGAATGCCATTGATTAGGA
Pt-infAfw	TCCTAATGCAATGTTTCGAG
Pt-infArev	TGAAGATTTAGCGCGAAGA
Pt-clpPfw	ATGCTGTTTGGATCGACGTA
Pt-clpPrev	TGCTTCTTGCGCTGACATA
Pt-chlNfw	TGGCAGAATTAGAAGAAGGA
Pt-chlNrev	CCAATAGGGAAAGGTGCT
Pt-chlLfw	AAGGTGGCATTGGTAAATCA
Pt-chlLrev	ACTGAAGCCGCAATACGA
Pt-chlBfw	AAGCCGCAGATAGAACTTTA
Pt-chlBrev	TGAAGCTGCGTGAGTTG
Pt-atpFfw	TTATTGGCCTATTGCTGG
Pt-atpFrev	ATCAATTACGCGTGAATGTA
Pt-atpEfw	CCTAATCACGCGCCACTTT
Pt-atpErev	CGTCCGCTTGCGATACTTTA
Pt-atpBfw	AGTGCGACAGATGGCTTGAT
Pt-atpBrev	GGCGGCTCATTCATTTGAC
Pt-atpAfw	CTGCGCCTGGTATTATTTCA
Pt-atpArev	TCCCCTGGATATGCTTCTCT
Pt-accDfw	TTTGAAGATAAACGGCGATT
Pt-accDrev	TCCACCGGAAGAACATACA
Pt-rpl33fw	TGGCTAAGAATAAAGATGTAAGG
Pt-rpl33rev	TCGAGTAGGTGTATTACGACG
Pt-rpl21fw	TTGAAACCGGAGGTGAACA
Pt-rpl21rev	CGAGCTGAATTTTGTCGATG
Pt-rpl20fw	ACGTGGGTATGTAGCTCGAA
Pt-rpl20rev	ACGCTATTTGTGCAAGCATT
Pt-rbcLfw	GCCCACCTCACGGTATTC
Pt-rbcLrev	GGTCAAGAACTGCGTGCA

2.4 Protein analysis

2.4.1 Determination of protein concentrations

Protein concentrations were determined using the Bio-Rad Protein Assay. Concentrations of distinct proteins were approximated by comparison to defined BSA amounts in Coomassie-stained polyacrylamide gels.

2.4.2 SDS polyacrylamide gel electrophoresis (SDS PAGE)

For protein analysis, 200 µl aliquots of *E. coli* cultures were pelleted and lysed in 40 µl 1x sample buffer. Following an incubation at 95°C for 5 min, samples were centrifuged and an appropriate aliquot of the supernatant was separated by SDS PAGE in a Mighty Small Vertical Electrophoresis Unit (Hoefer). A protein molecular weight marker (SM671, Fermentas) was run alongside samples. After electrophoresis, gels were subjected to Western blotting or Coomassie-staining.

4x sample buffer: 0,32 M Tris/HCl pH 6.8, 0,1 M EDTA, 0,4 M DTT, 8% (w/v) SDS, 4% (v/v) glycerol, 0,2 % (w/v) bromphenol blue

Acrylamide stock solution: Gel 30 (Roth)

Separating gel: 10% or 12% acrylamide, 375 mM Tris/HCl pH 8.8, 0,1% (w/v) SDS

Stacking gel: 4% acrylamide, 125 mM Tris/HCl pH 6.8, 0,1% (w/v) SDS

Electrophoresis buffer: 25 mM Tris, 192 mM glycin, 0,1% (w/v) SDS 0,1% (w/v)

Coomassie staining solution: Coomassie Brilliant Blue R250 in destain solution at 0,25 %

Destaining solution: isopropanol, 0,1 % (v/v) acetic acid

2.4.3 Immunoblotting

For immunostaining proteins were transferred onto nitrocellulose membranes (Hybond- C extra, Amersham Biosciences) in a Semi Dry Blot chamber (Bio-Rad). Following PAGE, gels were incubated in transfer buffer for 10 min. Membranes were blocked for 2 hours in TBST containing 5% (w/v) BSA at RT, incubated overnight with primary antibody diluted in TBST containing 5% (w/v) BSA, washed 3x 20 min in TBST, incubated with the secondary antibody diluted in TBST for 1 h, washed 3x 20 min in TBST and rinsed for 10 min in TBS. Before immunodetection using alkaline phosphatase was initiated the membranes were finally washed 3x 1 min in AP buffer and then exposed to AP sustrate solution. To stop the reaction membranes were washed subsequently in water.

Transfer buffer: 48 mM Tris, 39 mM glycin, 20% (v/v) methanol, 0,0375% (w/v)

TBS: SDS 10 mM Tris/HCl pH 7.5, 0,2% (w/v) SDS

TBST-BSA: 1% (w/v) BSA in TBST

TBST: 0,05 % (v/v) Tween 20 in TBS

AP buffer: 100 mM Tris/HCl pH 9.5, 100 mM NaCl, 5 mM MgCl2

AP substrate solution: 0,033% (w/v) NBT (nitroblue tetrazolium), 0,0165% (w/v) BCIP (5-bromo-4-chloro-3-indolyl phosphate) in AP buffer

Table 9: Antisera.

Antibody	Properties	Dilution	Reference
anti-His	Mouse IgG, raised against a polyhis-tagged protein	1:5000	Sigma
anti-AtpB	rabbit raised antibody against Zea mays AtpB	1:10000	Prof. Dr. A. Barkan, Oregon
anti-Atp1	raised against F1-ATPase subunit 1 of maize	1:10000	Prof. Dr. A. Fernie, MPI, Golm
anti-Aox1/2	rabbit polyclonal, raised against Aox1 c-terminus	1:40000	Agrisera
anti-Cox2	rabbit polyclonal, raised against synthetic peptides	1:40000	Agrisera
anti-Nad9	rabbit polyclonal, raised against wheat Nad9	1:5000	Prof. Dr. Bonnard, IBMP Strasbourg
Secondary	Anti-rabbit IgG-alkaline phosphatase conjugate	1:20000	Sigma
Secondary	Anti-mouse IgG-alkaline phosphatase conjugate	1:20000	Sigma

2.5 Recombinant protein expression

2.5.1 Plasmids for the expression of recombinant proteins

For both *Physcomitrella* RpoT expression constructs full-length ORFs were amplified with UTR primers PpT1cDNAp/PpT1cDNAm and PpT2cDNAp/PpT2cDNAm. PCR products were cloned into pDrive and verified by boarder sequencing using vector primers M13-seq-F and M13-seq-R. To amplify the full-length *RpoT* coding frames, excluding the 5' portions encoding the predicted transit peptides, verified pDrive clones and primers with restriction enzyme overhangs were used (PpRpoTmp1: ExPp1pNde1/ExPp1mSal1, PpRpoTmp2: ExPp2pKpn1/ExPp2mSal1). Sequences encoding amino acids 88-1087 of PpRpoTmp1 and amino acids 73-1065 of PpRpoTmp2 were amplified and ligated into pColdI (TaKaRa) to yield plasmids pCold-HisRpoTmp1 and pCold-HisRpoTmp2 encoding hexahistidine-tagged proteins. The quality and integrity of HisRpoTmp1 and HisRpoTmp2 coding-frames was verified by sequencing using primers of the coding sequences reported before (Richter *et al.* 2002).

Table 10: Oligonucleotides for generation of expression constructs

Oligo Name	**Oligo Sequence (5‘ – 3‘)**
PpT1cDNAp	TTGGTTGGGACTCCGCGTCG
PpT1cDNAm	CGCTGCTCCACCAGTCCTGC
PpT2cDNAp	ATGGTGGCGTGTTGCTTCGG
PpT2cDNAm	AGCAAACACTGCCGGTGAAA
ExPp1pNde1	tcatatgGCGGCGGCTGCTGTT
ExPp1mSal1	tgtcgacCTAATTAAAGAAATAAGGTGCACG
ExPp2pKpn1	tggtaccACAGCGCATCAGGCCTC

ExPp2mSal1	tgtcgacCTAGTTGAAAAAATATGGAGATTTGAGAAC

Lowercase letters correspond to non-annealing nucleotides for the introduction of restriction sites.

2.5.2 Protein expression in *E. coli*

PpRpoTmp1 and PpRpoTmp2 were overexpressed from pCold-HisRpoTmp1 and pCold-HisRpoTmp2 in an *E. coli* BL21 Codon Plus RIL strain (Stratagene). 250 ml fresh LB medium containing 100 µg/ml Ampicillin were inoculated with 3,5 ml of an overnight culture grown under standard conditions in LB medium and supplemented with 100 µg/ml Ampicillin and 30 µg/ml Chloramphenicol. 250 ml cultures were grown under standard conditions until an OD_{600} of 0,5 to 0,6. Cultures were refrigerated at 15°C. After 30 minutes IPTG was added to a final concentration of 0,9 mM, and the cells were cultured for 21 hours at 15°C until harvested by centrifugation (10 min, 6000xg, 4°C) in a Megafuge 1.0 R (Heraeus). At 19 hours 250 µl aliquots of cultures were pelleted and subjected to SDS-PAGE analysis to confirm recombinant protein expression.

2.5.3 Purification of His-tagged recombinant protein from *E. coli*

Recombinant RpoTmp1 and RpoTmp2 was prepared from 250 ml of cell culture. Harvested cells (10 min at 4000 g and 4°C) were resuspended in 150 ml buffer I, recentrifuged and resuspended in 5 ml buffer II. Cells were broken by shaking in a bead mill (Retsch) with 1 g glass beads (⌀ 0.17-0.18 mm) per 0,75 ml cells at maximum frequency (10 min, 4°C). The cell lysate was cleared by centrifugation (10 min, 6000xg, 4°C) and transferred into a fresh 15-ml polypropylene tube. The left pellet was again shaken in bead mill after addition of 5 ml buffer II. Soluble fractions obtained from both extraction steps were pooled and recentrifuged (10 min, 6000xg, 4°C). The lysate and 0.5 ml (bed volume) buffer I equilibrated Ni2+- NTA-agarose (Qiagen) were distributed over two 15-ml polypropylene tubes and incubated in a spinning wheel (16 h, 4°C) to allow proteins to bind to the matrix. Subsequently, matrix with bound proteins was pelleted (2 min, 1000xg, 4°C) and the supernatant was removed. The matrix was then washed in 6 ml buffer III, recentrifuged (2 min, 1000xg, 4°C) and resuspended in 3 ml buffer III. The slurry was transferred into a 1-ml polypropylene column (Qiagen) and left to settle. The column was washed with 3x 3 ml buffer IV and 3x 1 ml buffer V. 3x 0,5 ml buffer VI was used to elute proteins from the matrix. The volume of the eluate was brought down to 300 µl by centrifugation (9000xg, 4°C) in 50000-MWCO-*Centricon* centrifugal filter devices (Millipore). Finally, protein samples were dialyzed for 16 h at 4°C against buffer VII and dialyzed proteins stored in aliquots at -20°C for use within six month. Protein preparations were analysed by SDS-PAGE and immunostaining with a polyhistidine antibody (*anti*-His, Mouse IgG, 1:5000, Sigma).

Buffer I: 100 mM Tris/HCl, pH 7.8; 300 mM NaCl; 5 mM imidazole

Buffer II: Buffer I with 1 mM PMSF, 1 mM benzamidine, 0.5 mM DTT

Buffer III: 20 mM Tris/HCl, pH 7.0; 300 mM NaCl; 5 mM imidazole

Buffer IV: 20 mM Tris/HCl pH 7.0; 300 mM NaCl; 10 mM imidazole

Buffer V: 20 mM Tris/HCl pH 7.0; 300 mM NaCl; 30 mM imidazole

Buffer VI: 20 mM Tris/HCl pH 7.0; 300 mM NaCl; 100 mM imidazole

Buffer VII: 20 mM Tris/HCl pH 7.8; 100 mM NaCl, 0,5 mM EDTA, 1 mM DTT, 50% (v/v) glycerol

2.6 *In vitro* transcription

2.6.1 Template construction

Organellar DNA fragments were PCR-amplified from total DNA with primer pairs listed in Table 11. PCR products were ligated into pKL23 (Liere and Maliga 1999) upstream of described transcriptional terminators. Plasmids were purified employing the Plasmid Midi Kit (Qiagen).

Table 11: Construction of in vitro transcription templates.

Oligo Name	Oligo Sequence (5‘ – 3‘)
atp1KL+SacI	gcagagctcGCCAATGAAGTAGCGGAGGCCC
atp1KL-EcoRI	gcagaattcAAGAATTTTGGGGCATTTCTATATT
trnEKL+SacI	gcagagctcCTACGTGTCATGTAGGCTGC
trnEKL-EcoRI	gcagaattcCGTGTTATAATGCATCCAGATCACG
tmetKL+Sac1	gcagagctcTCCGGGCACTATGGTAAGAC
tmetKL2-EcoR1	gcagaattcCATTCAGATTATGAGCCTGACG
26SKL1+Sac1	gcagagctcGGTGTGATAAGGCGAATACTCA
26SKL1-EcoR1	gcagaattcAGCGTCCTTCCTTCTCAATGC
atp6KL+SacI	gcagagctcTCGCTGCGCGAATGTAGGGC
atp6KL-EcoRI	gcagaattcATTGCAAATTGTTCTAGCGGAC
atpBKL+Sac1	gcagagctcTTCTACATTTGATGCAACTCAGATT
atpBKL-EcoR1	gcagaattcAAGCTCCAAAAGTACGGGAAT
psbAKL+Sac1	gcagagctcAATCATCCACTTTCATCCG
psbAKL-EcoR1	gcagaattcGCAGTCATGGTAAATCTTGG

Lowercase letters correspond to non-annealing nucleotides for the introduction of restriction sites.

2.6.2 *In vitro* transcription assay

In vitro transcription assays were carried out for 45 min at 30°C and followed the protocol of (Kühn *et al.* 2007). Briefly, *in vitro* transcription reactions (15 µl) were set up as follows: 6.7 mM Tris/HCl (pH 7.9), 6.7 mM KCl, 6.7 mM MgCl2, 0.67 mM DTT, 0.067% (w/v) BSA, 267 µM each ATP, CTP and GTP, 13 µM unlabelled UTP and 10 µCi of [α-32P]-UTP (3000 Ci/mmol), 24 U RNase inhibitor and 200 ng of pKL23 template DNA. Transcription reactions were initiated by the addition of 400 fmol recombinant RpoTmp1 or RpoTmp2 and stopped by adding 115 µl RNA extraction buffer (6 M urea, 360 mM NaCl, 20 mM EDTA, 10 mM TRis/HCl pH 8.0, 1% (w/v) SDS). 20 µl 2.25 M sodium acetate (pH 5.2) were added to each reaction and RNA was extracted by phenol/chloroform/isoamyl alcohol (25:24:1). After ethanol precipitation and washing with 70% (v/v) ethanol, formamide buffer (95% (v/v)

formamide; 0,02% (w/v) bromphenol blue; 0,02% (w/v) xylene cyanol) was added to the dried RNA pellets and resolved by denaturing PAGE.

2.6.3 5'-end mapping of *in vitro*-synthesised RNAs

5' ends of *in vitro*-synthesised transcripts were mapped after transcription assays were carried out as described in 2.6.2 using unlabelled UTP. After purification, samples were dissolved in ultrapure water and TAP-treated. Subsequently, 5'-RACE was performed as described in 2.3.7. Notably, 5'-RACE on non-ligated samples (no ligase) was used as a control. Reverse primers P2hisa and P3hisa annealing to the *hisa* terminator sequence of pKL23 derived transcript portions (see Suppl. Fig. 7) were used for cDNA synthesis and PCR, respectively. RUMSH2 served as the forward primer in PCR reactions. PCR products were cloned into pDRIVE and obtained clones sequenced with pDRIVE border primers M13forward and M13reverse.

Table 12: Primers used for 5'-end mapping of in vitro-synthesised RNAs

Oligo Name	Oligo Sequence (5' – 3')
P2hisa	CACATCGCCTGAAAGACT
P3hisa	GGATGATGGTGATGATGGTGG

2.7 *PpRpoT3* cDNA cloning

The *blast* analysis of the published *Physcomitrella* whole genome sequence (Rensing *et al.* 2008) revealed a third genomic locus with strong similarity to plant phage-type RNA polymerases on PHYPAscaffold_241 (NW_001865497). Putative coding regions were used to generate primers for the amplification of cDNA ends. 3'- and 5'- RACE reactions employed the CapFishing Full-length cDNA Premix kit (Seegene) and Phusion hot start DNA polymerase according to the protocols of the manufacturers. For the cDNA synthesis, 1 µg of polyA-enriched *Physcomitrella* protonema RNA was reverse transcribed using the SuperScriptTM III RNase H- reverse transcriptase kit. Notably, 5'RACE reverse transcription employed two gene-specific primers annealing to the *in-silico* predicted lower exon 1 region (5racePpT3gsp1/ 5racePpT3gsp2). While 3'- and 5'- RACE adaptor primers were provided in the kit, gene-specific primers (gsp) were employed for two rounds of PCR to enhance the specificity of the reaction (see Tab. 13). To verify the cloned transcript-ends a primer walking strategy was used until no specific amplification products were obtained after a second round of 35 PCR cycles (5racePpT3gsp7 and 3racePpT3gsp4). After the identification of both *PpRpoT3* transcript termini the whole *PpRpoT3* coding region was amplified using primers PpT3fw1 and PpT3m3169 and the fragment was cloned into pDrive (pDriveRpoT3cDNA). To map the *PpRpoT3* gene structure subsequent sequencing of the whole PpRpoT3 cDNA employed primers used in previous steps of the cloning procedure and sequencing primers shown in Table 13.

Table 13: Oligonucleotides for *PpRpoT3* cDNA cloning

Oligo Name	Oligo Sequence (5‘ – 3‘)
5racePpT3gsp1	CCAACTGAGACGGAGGCAAC
5racePpT3gsp2	GCACTACGGTAAGGATTCAAAGACC
5racePpT3gsp3	AGTTTGCACTTCCGCTGTCG
5racePpT3gsp4	AACTTGCGCAATCCTTGACC
5racePpT3gsp5	GACCTGCTGCCTCCACAATG
5racePpT3gsp6	CCTCCACAATGCAACAATTCG
5racePpT3gsp7	AAGCAGAAATAAGAGAGAGAGGGAGC
3racePpT3gsp1	GGAGTCCTGGACATCAGGGAGG
3racePpT3gsp2	TGGACATCAGGGAGGTTCTTAAAGCCC
3racePpT3gsp3	TGCTCTCTTTGCAACTGGGTGACG
3racePpT3gsp4	TTCAGACCAACGTCGGAATGG
PpT3p43	CATTGTGGAGGCAGCAGGTC
PpT3m3169	TCCATTCGTCACCCAGTTGC
Seq4T3fw	ACGAGTTCCGAGCGACGATCC
Seq3T3rev	ACCAAAGGATTGCACTCCACAACACC
Seq2T3fw	TGCAACATTATGCAGCCCTTGG
Seq1T3rev	CGCTGCATGTTTACTGGCCTCGG
PpT3exon2	GCCTTTCCTCTTCGTTTCGT

2.8 GFP targeting studies

2.8.1 GFP targeting constructs

To generate the pPpRpoT3-GFP constructs and to drive the expression of the fusion protein PpRpoT3-GFP, a PpRpoT3 fragment containing the full 5'UTR and encoding 78 N-terminal amino acids of PpRpoT3 was amplified from pDriveRpoT3cDNA using primers PpT3gfpfwXbaI/PpT3gfprevSalI. PCR products were *Xba*I/*Sal*I-digested and ligated into the *Spe*I/*Sal*I-cleaved vector pOL−GFP S65C (Peeters *et al.* 2000). Mitochondrial CoxIV-GFP and plastidial RecA-GFP (Peeters *et al.* 2000) control constructs were kindly provided by I. Small (UWA, Perth, Australia). PpT3gfpXba1-2 was used for amplification of a construct lacking all but 12 nt of the 5'UTR.

Table 14: Oligonucleotides for *PpRpoT3*-gfp fusion constructs

Oligo Name	Oligo Sequence (5‘ – 3‘)
PpT3gfpXba1	cactctagaCTGCTTGCGTTGCTTTGC
PpT3gfpSalI	atagtcgacCAAGGATGTCTTCCAGAGGTG
PpT3gfpXba1-2	cactctagaGAGTTGAATACT**ATG**TGGC

Lowercase letters correspond to non-annealing nucleotides for the introduction of restriction sites.

2.8.2 Transient expression in *Arabidopsis* protoplasts and microscopy

Protoplasts were isolated from *Arabidopsis* leaves grown for 4 weeks under long day conditions (23°C, 16/8 hr light/dark) as described before (Yoo *et al.* 2007). Briefly, 100 µl protoplasts (2×10^6 protoplasts/ml) were transfected with 20 µg pOL-PpRpoT3gfp plasmid in

40% polyethylene glycol 4000, 0,8 M mannitol and 1 mM CaCl2. Transformed protoplasts were kept in the dark and examined 48h after transfection by confocal laser scanning microscopy with a Leica TCS SP2 using 488 nm excitation and two-channel measurement of emission from 510 to 580 nm (green/GFP) and >590 nm (red/chlorophyll).

2.9 Alignments and phylogeny

In general sequence information was retrieved from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/) employing the BLASTP and TBLASTN algorithms and from the Joint Genome Institute (http://genome.jgi-psf.org/) if not indicated otherwise. For initial selection of sequences to be included in the phylogenetic analysis COBALT, a **Co**nstraint-**b**ased Multiple **Al**ignment **T**ool (Papadopoulos and Agarwala 2007) was used. For translation and alignment, sequences were subsequently imported into GENEIOUS (Drummond *et al.* 2010, Drummond *et al.* 2008). Multiple protein sequence alignment were generated using CLUSTALW (Thompson *et al.* 1994) implemented in the GENEIOUS package using GONNET cost matrix with gap open cost set to 10 and gap extend cost set to 0,1 as the default setting. All alignments were refined manually and poorly aligned regions were generally excluded from following analyses. For a phylogenetic reconstruction based on Bayesian statistics, the MRBAYES program version 3.1 (Ronquist and Huelsenbeck 2003) was used. In addition, maximum likelihood (phyML; (Guindon and Gascuel 2003)) and neighbour joining analysis (Saitou and Nei 1987) all implemented in the GENEIOUS package were conducted.

2.10 Material

Chemicals and biochemicals were generally purchased from Roth, ICN Biomedical, Serva, Sigma or Becton-Dickinson. Radiochemicals were provided by Amersham Buchler and PerkinElmer. Deoxyribonucleoside triphosphates and ribonucleoside triphosphates were obtained from Fermentas. Oligonucleotides were purchased from Metabion, Sigma or Eurogentec. Ultrapure water was taken from a USF Purelab Plus system. Other materials are specified in the previous sections.

2.11 Providers

Applied Biosystems	Applied Biosystems, Weiterstadt, Germany
Ambion	Ambion, Inc., Austin, TX, USA
Amersham Biosciences	Amersham Biosciences Europe GmbH, Freiburg, Germany
Biometra	Biometra GmbH, Göttingen, Germany
Bio-Rad	Bio-Rad Laboratories, Richmond, VA, USA
Biozym	Biozym Diagnostik GmbH, Hameln, Germany
Braun	Braun GmbH, Kronberg, Germany
Calbiochem	Calbiochem Merck Biosciences GmbH, Schwalbach, Germany
Clontech	Clontech Takara Bio Company, Takara Bio Inc., Otsu, Japan
DuPont	DuPont de Nemours GmbH, Bad Homburg, Germany
Epicentre	Epicentre Biotechnologies, Madison, WI, USA

Eurogentec	Eurogentec, Seraing, Belgium
Fermentas	Fermentas GmbH, St. Leon-Rot, Germany
Finnzyme	Finnzyme Oy, Espoo, Finland
GE Healthcare	GE Healthcare Europe GmbH, Freiburg, Germany
Heraeus	Heraeus, Hanau, Germany
Invitrogen	Invitrogen GmbH, Karlsruhe, Germany
Macherey-Nagel	Macherey-Nagel, Düren, Germany
Macrogen	Macrogen Inc., Seoul, Korea
Metabion	Metabion international AG, Martinsried, Germany
Millipore	Millipore Corp., Bedford, USA
MP Biomedical	MP Biomedical, Illkirch, France
Nalgene	Nalgene Labware, Rochester, NY, USA
Operon	Operon Biotechnologies GmbH, Köln, Germany
peqLab	peqLab Biotechnologie GmbH, Erlangen, Germany
Perkin Elmer	Perkin Elmer LAS (Germany) GmbH, Rodgau, Germany
Pierce	Pierce, Rockford, IL, USA
Promega	Promega Corp., Madison, WI, USA
Q-Biogene	Q-Biogene, Heidelberg, Germany
Qiagen	Qiagen, Hilden, Germany
Retsch	Retsch GmbH, Haan, Germany
Roche	Roche Diagnostics GmbH, Mannheim, Germany
Roth	Carl Roth GmbH & Co. KG, Karlsruhe, Germany
Seegene	Seegene, Rockville, USA
Serva	Serva Feinbiochemika, Heidelberg, Germany
Sigma	Sigma Chemical Company, St. Luis, MO, USA
Sorvall	Kendro Laboratory Products GmbH, Langenselbold, Germany
Stratagene	Stratagene, La Jolla, CA, USA
Takara	Takara Bio Inc., Otsu, Japan
Thermo Scientific	Thermo Scientific LED GmbH, Langenselbold, Germany
USF	USF, Seral Reinstwassersysteme GmbH, Germany
Whatman	Whatman Paper, Maidstone, UK
Yakult	Yakult, Tokyo, Japan

3 RESULTS

3.1 Analysis of deviating functions of organellar phage-type RNA polymerases - I. *Physcomitrella patens*

3.1.1 Targeted gene knock-out for two phage-type RNA polymerases in *P. patens*

When knock-out experiments were initiated only two clearly divergent nuclear RpoT genes were identified in *P. patens* (Kabeya *et al.* 2002, Richter 2003, Richter *et al.* 2002) coding for the phage-type RNA polymerases PpRpoTmp1 and PpRpoTmp2. The two genes encode N-terminal transit peptides for dual targeting to both mitochondria and chloroplasts (see 1.2.6). Therefore potentially two distinct RpoT activities were thought to reside in both organelles, a situation also observable in *Arabidopsis thaliana* (see 1.2.6). Taking advantage of *Physcomitrella* being amenable to targeted gene knock-out by homologous recombination (Hohe *et al.* 2004, Schaefer and Zryd 1997) an attempt was made to establish single knock-out mutants for both genes in order to elucidate potentially distinct functions of the two enzymes.

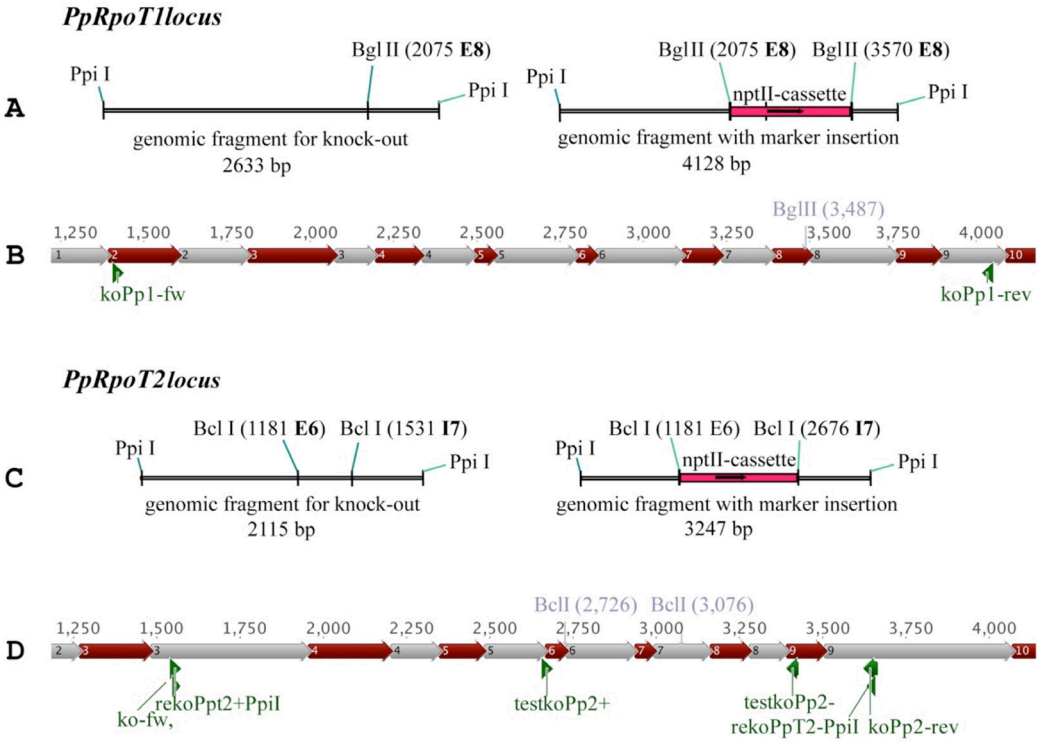


Figure 5: *PpRpoTmp1* and *PpRpoTmp2* genomic loci and cloning strategy for direct gene knock-out. In A and C insertion sites of the *npt*II cassette are depicted. B and D illustrate the genomic loci of both genes with the respective intron-exon structure. Numbered exons are depicted in red, introns in gray. Primers used for amplification of knock-out cassettes and for testing of transformed moss plants are in green.

Geneticin resistance (35S:*npt*II) cassettes were used to disrupt PpRpoTmp1 and PpRpoTmp2 function (see Fig.5A and 5C). For *PpRpoTmp2* insertion constructs were designed in such a way as to enable double crossovers on either side of the antibiotic resistance cassette which would result in a replacement of the genomic region from exon 6 to intron 7 with the 35S:*npt*II cassette (see Fig.5D and 6). For *PpRpoTmp1* the 35S:*npt*II cassette was inserted via a single BglII site, resulting in disruption of the *PpRpoTmp1* open reading frame in exon 8 (see Fig.5B). *Physcomitrella* wild-type plants were transformed with linearised DNA from these constructs to generate the single gene mutants, *PprpoTmp1* and *PprpoTmp2*.

After protoplast transformation and selection/regeneration resistant *Physcomitrella* plants were analysed by PCR to identify mutant lines with integration of the disruption constructs at the *PpRpoTmp1 or PpRpoTmp2* genomic loci. For *PpRpoTmp1* no homoplasmic moss lines were obtained. Instead, transformants were found to be 2n and did possess both the wild-type and the *PprpoTmp1-nptII* allele (personal communication Dr. Decker, Freiburg). This in turn suggests that *PpRpoTmp1* has an indispensable function in *Physcomitrella patens* and disruption of the *PpRpoTmp1* locus is likely lethal.

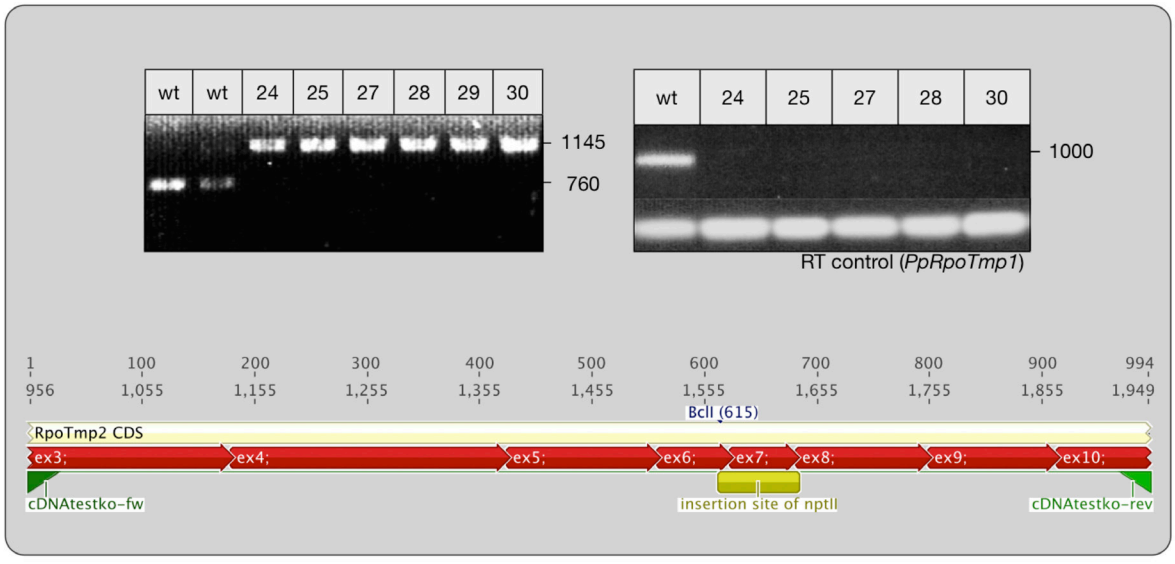


Figure 6: Analysis of Pprpotmp2 knock-out lines. In the upper left panel genomic PCR results confirm integration of *npt*II marker cassette (genomic locus is illustrated in Figure 5C and 5D). The RT-PCR results as shown in the upper right panel are indicative of disruption of the PpRpoTmp2 coding frame, with RT-PCR control reactions using *PpRpoTmp1* specific primers below. Binding sites of primers used for PCR after random primed cDNA synthesis and the *npt*II insertion/deletion site are indicated in the cartoon.

For *RpoTmp2* six independent knock-out lines (ko24, ko25, ko27, ko28, ko29 and ko30) from 34 plants (<18%) were obtained proven by sequencing of genomic PCR products (Fig. 6 left panel) and RT-PCR (Fig. 6 right panel). The *PpRpoTmp2* locus was disrupted and all recombined alleles were found to be null, implying that PpRpoTmp2 has a dispensable function in the moss or alternatively, that its function can be complemented at least in part by another RNA polymerase.

3.1.2 Knock-out of PpRpoTmp2 causes developmental and strong phenotypical aberrations

On minimal solid medium (Knop) formation of a *Physcomitrella* colony starts from a small cluster of cells placed on the plates. Filamentous protonema cells start to grow and establish a network of filaments. Subsequently, gametophores resembling a small plant with leaf-like structures called leaflets and root-like rhizoids develop from a fraction of the protonema. To investigate the influence of the *RpoTmp2* gene disruption on *P. patens* cell growth and development, *PprpoTmp2* colonies formed on Knop medium as well as protonema from liquid culture were compared with wild-type moss plants (Fig. 7).

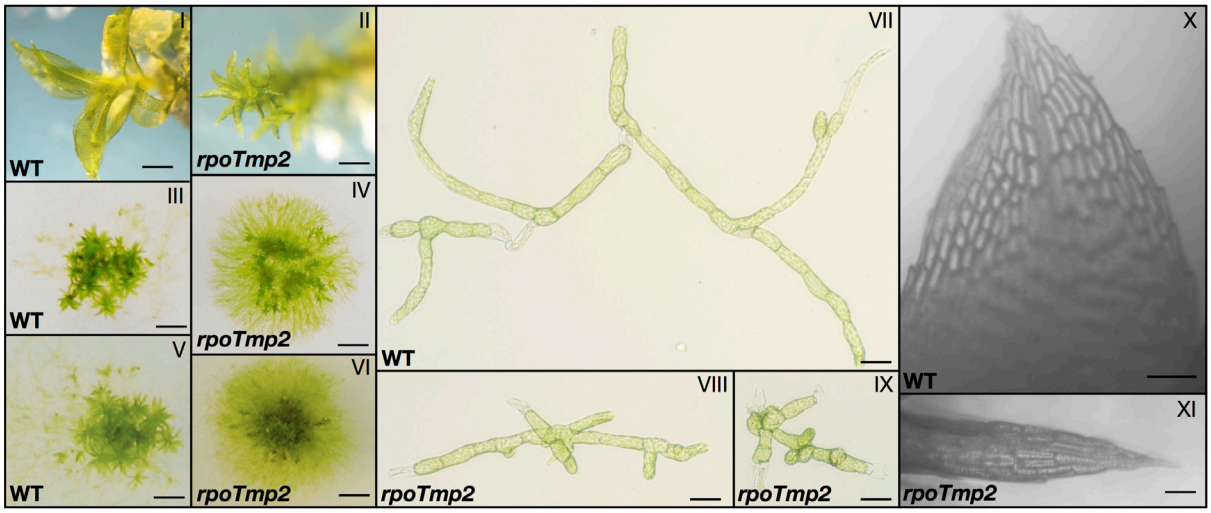


Figure 7: Growth and development of wild-type and *Pprpotmp2* insertional disruptant moss plants. I-VI: Morphology of colonies grown under standard conditions on solid Knop plates (I-II), under low light (20µE, III-IV) and high light (200µE, V-VI) for 4 weeks. VII-IX: Protonemata from the wild-type background and of the mutant cultivated in liquid Knop medium for 18 days. Leaflets of wild type (X) and *Pprpotmp2* disruptant (XI) plants taken from colonies shown in I and II. Bars = 1 mm (I-II), 7 mm (III-VI), 80 µm (VII-IX) and 200 µm (X-XI).

Since RpoT proteins in plants are known to be involved in organelle transcription (see 1.2.6), potential mitochondrial and/or plastid phenotypical aberrations were of primary interest. Light microscopy revealed no abnormalities of chloroplast shape or number for leaflets and protonema tissue in comparison to the WT (Fig. 9). To examine mitochondria and chloroplast fine structure, transmission electron microscopy (TEM) was carried out, but revealed no notable differences between WT and *PprpoTmp2* plants (Fig. 8).

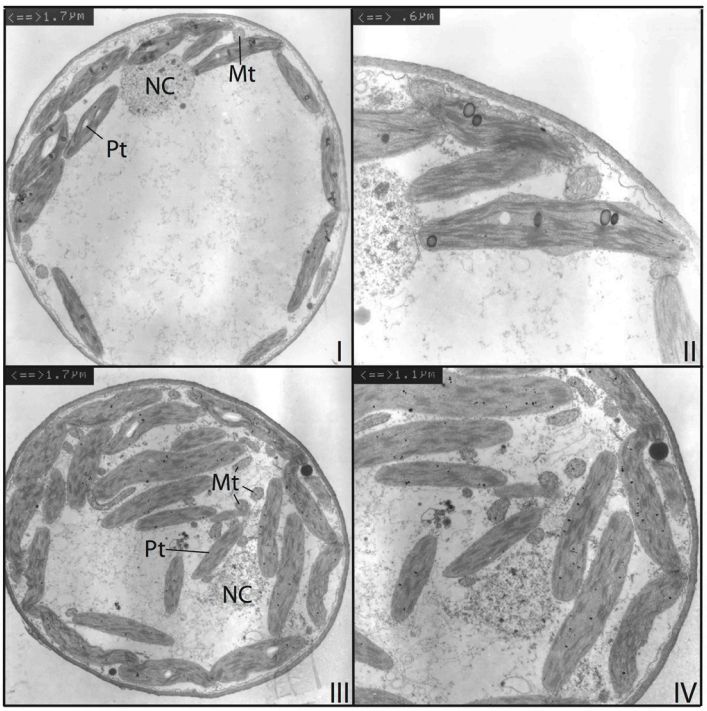


Figure 8: Transmission electron microscopy (TEM) of wild-type (I-II) and *Pprpotmp2* (III-IV) protonema cells. Two wild-type and three knock out line (lines 25, 27 and 29) cultures were analysed. No differences in the fine structure of mitochondria (Mt) or plastids (Pt) from mutant plants are observable.

To the contrary after four weeks of growth on solid Knop medium under standard conditions, gametophores of all *PprpoTmp2* knock out lines were clearly distinguishable from the WT (Fig. 7). Mutant lines with wrinkly javelin-like leaflets were strongly reduced in size (Fig. 7 and 9) and appeared slightly darker compared to wild-type plants. While pulse-amplitude modulated (PAM) chlorophyll fluorescence assays were not indicative of changes in photosynthetic performance of the mutant (data not shown), differences in the relative pigment contents were observed for violaxanthin and betacarotene. Both photo-protective pigments were slightly elevated in *PprpoTmp2* plants (Tab. 15).

Table 15: Relative pigment contents were compared between wild-type and mutant plants (line 25, 27 and 29) grown under standard conditions.

	Chla+Chlb	Neoxanthin/Chla	Violaxanthin/Chla	Lutein/Chla	BetaCar/Chla	Chla/Chlb
WT	41,20	85,70	0,04	221,40	64,70	2,35
PprpoTmp2	43,78	88,83	0,07	230,43	80,15	2,35
% PprpoTmp2	106	104	156	104	124	100

While single protonema cells were not significantly reduced in size, the length of whole filaments was considerably reduced in mutant plants growing on Knop-plates as well as from liquid cultures. Two weeks after protoplastation and growth in liquid culture under standard conditions, the majority of *PprpoTmp2* plants were found to be in four-cell state, while wild-type filaments where beyond eight cell stage and did show the typical branching pattern. Interestingly the mutant filaments never grew to normal length, but started to branch before reaching the length of wild-type protonema (Fig. 7). To substantiate the conclusion that the observed phenotypical aberrations are indeed the direct result of a lacking PpRpoTmp2 activity, the wild-type genotype was reconstituted in mutant plants. Transformation with a wild-type 2115 bp genomic fragment (see Fig. 5c) did indeed result in a complete rescue of the *PprpoTmp2* phenotype (personal communication Dr. Decker, Freiburg).

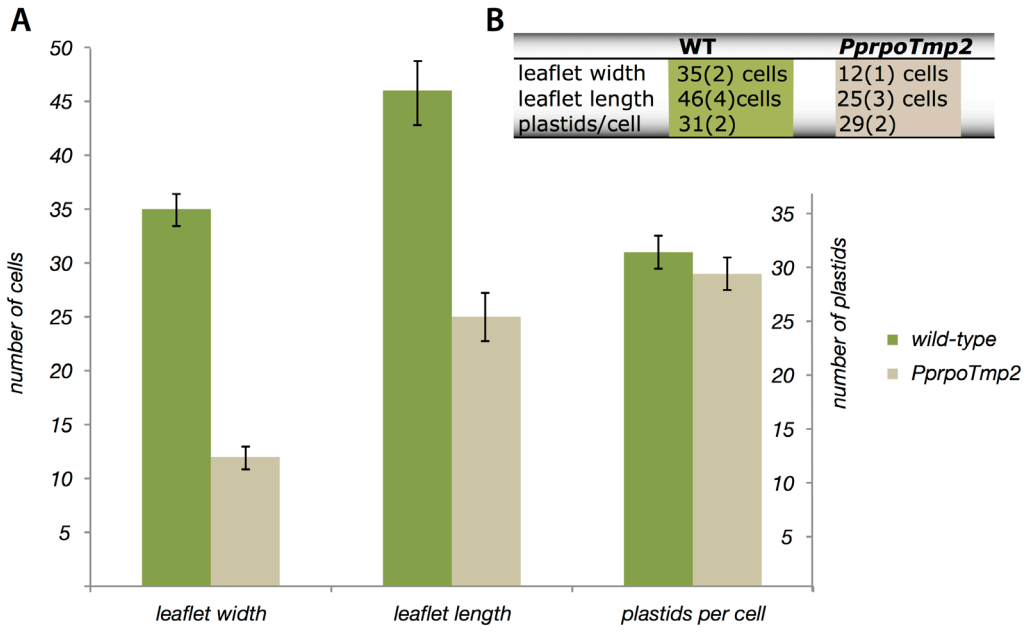


Figure 9: Microscopic analysis of phenotypical aberrations of *PprpoTmp2* gametophores. To count the number of plastids per cell 10 cells per leaflet from five wild-type plants and four *PprpoTmp2* plants (line ko24, ko27, ko28, ko29) were randomly chosen. The number of cells in width and length was determined from 10 wild-type and 8 mutant plants (line ko24, ko27, ko28, ko29) counting 2 leaflets each. Standard deviation is indicated in brackets.

3.1.3 Organellar steady-state transcript levels in *rpoTmp2*

The aberrations in leaflet morphology displayed by *PprpoTmp2* plants are reminiscent of leaf phenotypes described for mutants impaired in mitochondrial function (de Longevialle *et al.* 2007, Reichheld *et al.* 2007, Van Aken *et al.* 2007). Furthermore, the developmental delay throughout all stages of the life-cycle observed in *Physcomitrella rpotmp2* plants is also indicative of a mitochondrial defect (Zsigmond *et al.* 2008). This interpretation is substantiated by the recent analysis of a developmentally and growth retarded *Physcomitrella* RECA1 mutant in which mitochondrial genome stability is disrupted (Odahara *et al.* 2009).

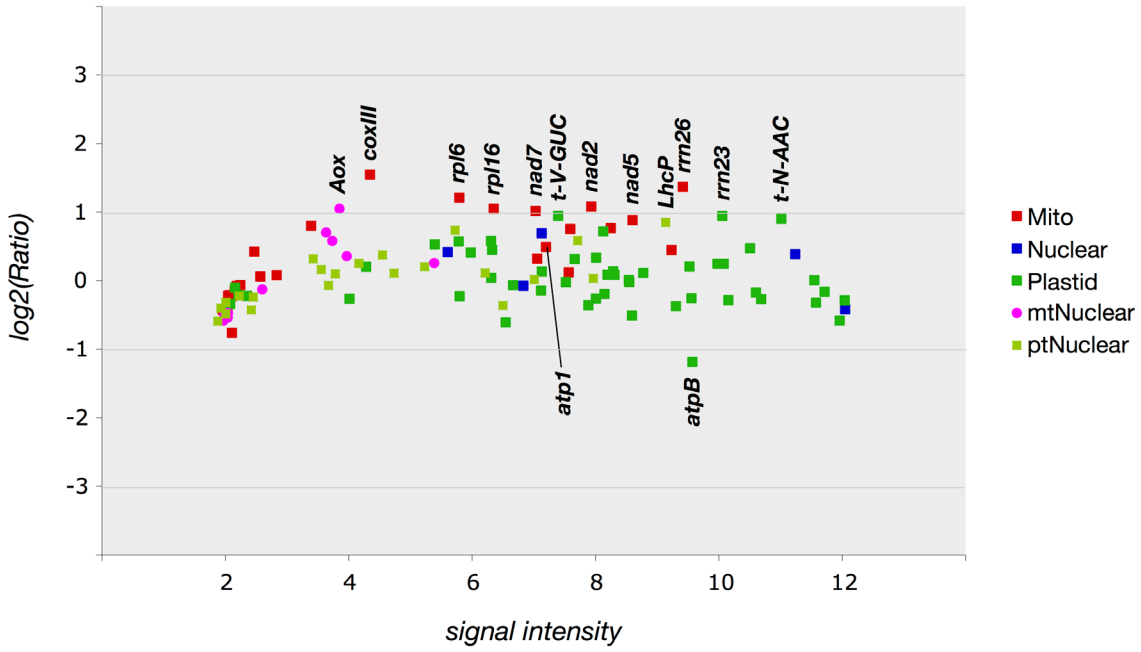


Figure 10: Macro-array analysis of steady-state transcript levels. Transcript levels are depicted as the log2 ratio of transcript levels in mutants compared with levels in wild-type moss plants. Transcript abundances of all genes were determined in both mutant and wild-type plants at the protonema stage (10 days of growth). Four biological replicates of the wild-type and six biological replicates of the *PprpoTmp2* genotype were averaged; two technical replicates were spotted per membrane. The nuclear 18S rRNA gene was used for data normalisation.

On the other hand, an *Arabidopsis* line lacking AtRpoTmp did show delayed development, which was most pronounced during juvenile stage. Furthermore, this mutant was delayed in flowering time and had short roots and an altered leaf shape; however, no changes in the accumulation of mitochondrial transcripts tested were found. Instead, the induction of several plastid genes in dark-grown seedlings upon illumination was reported to be delayed (Baba *et al.* 2004). Thus, molecular defects in plastid transcription were suggested to account for the observed phenotype. RpoTmp2 is a nuclear-encoded organellar RNAP. However, the requirement of the PpRpoTmp2 protein *in vivo* for mitochondrial or plastid transcription is not clear especially in view of the partial contradictions in localisation studies (see 1.2.6) and the lack of highly specific antibodies for RpoTmp1 and RpoTmp2. Therefore the accumulation of mitochondrial and plastid transcripts from *rpoTmp2* protonema cultures (line 27, 28, 29) was investigated using dot-blot analysis of 45 plastid and 22 mitochondrial genes. Furthermore, the macro-array included 27 nuclear encoded genes of which 16 encode plastid and 6 mitochondrial proteins. Since the mitochondrial genome of *Physcomitrella patens* was sequenced only recently (Terasawa *et al.* 2007), it was necessary to deduce primers from the sequenced *Marchantia polymorpha* chondrome (Oda *et al.* 1992). In general only marginal differences were seen when comparing normalised plastidial and mitochondrial transcript levels between *rpoTmp2* and the wild type. Only one transcript, the plastid *atpB*, was reduced more than two-fold, whereas six transcripts were elevated more than two-fold (see Fig. 10), notably all comprising genes encoding mitochondrial gene products. In general, all mitochondrial transcripts seem to be elevated, though to different extents. To further substantiate those results quantitative real time RT-PCR (qRT-PCR) was facilitated and additional mitochondrial transcripts were analysed when the mitochondrial genome sequence was published (Terasawa *et al.* 2007).

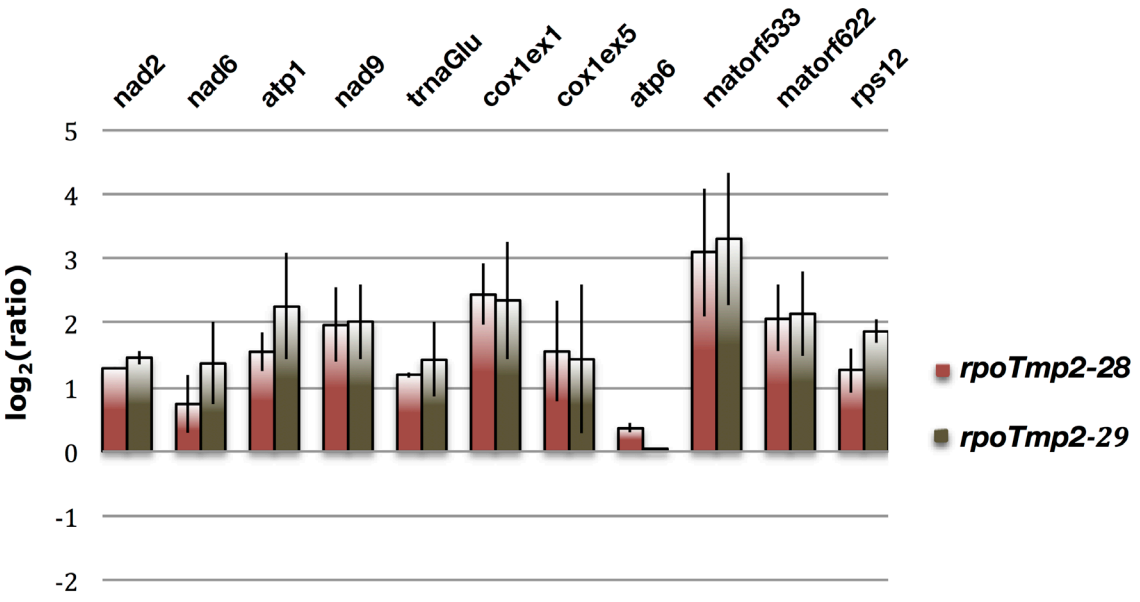


Figure 11: Mitochondrial Steady-state Transcript Levels in *PprpoTmp2*. Transcript levels are depicted as the log2 ratio of transcript levels in mutants compared to the wild-type plants. Transcript abundances were determined by qRT-PCR in both mutant and wild-type plants at the protonema stage. Three technical replicates were averaged per genotype; standard errors are indicated. The nuclear 18S rRNA transcripts were used for data normalisation.

The results for *nad2*, *nad6* and *atp1* as shown in Figure 11 clearly demonstrate that the observation of elevated mitochondrial steady-state transcripts is reproducible by real time RT-PCR. Moreover the elevation of mitochondrial steady-state transcripts in *rpoTmp2* measured by quantitative real time RT-PCR appears more pronounced, even though in both experiments data were normalised to the steady-state transcript levels of the nuclear encoded 18S ribosomal RNA. While macro-array results exhibit an 1,5-fold elevation for *atp1* transcripts in the mutant, the ratio in quantitative real time RT-PCR is 3-fold for *rpoTmp2*-28 and more than 4-fold for *rpoTmp2*-29 (see Fig. 11). All steady-state transcript levels tested by quantitative real time RT-PCR were found to be enhanced in the mutant but for *atp6*.

Strand specific Northern analysis for the mitochondrial genes *rrn18*, *rpl5* and *rps1*, as well as for the plastid *rpoB* gene were not indicative of changes in the accumulation of their corresponding steady-state transcript levels in the mutant (Fig. 12B). To the contrary steady-state transcript levels of *cox2* are elevated in *PprpoTmp2* plants and this elevation appears to be reflected in the accumulation of steady-state cox2 protein levels.

The reduction of *atpB* steady-state transcript levels as seen in the dot-blot experiments was not found to have any effect on the steady-state protein level, since no difference in the atpB protein accumulation was observable in Western analysis (Fig. 12A bottom). Moreover, strand specific Northern data are in favour of a lesser or no reduction of the sense transcript of *atpB* (Fig. 12 A top).

In the nuclear genome of *Physcomitrella patens* three putative open reading frames are found with high similarity to the *A*lternative *Ox*idase proteins found in flowering plants. Studies on Aox have revealed that its expression is induced by a variety of treatments often referred to as stresses, such as the use of inhibitors of the mitochondrial electron transport chain (Clifton *et al.* 2005, Van Aken *et al.* 2009) or in plants with mutations of respiratory complex subunits

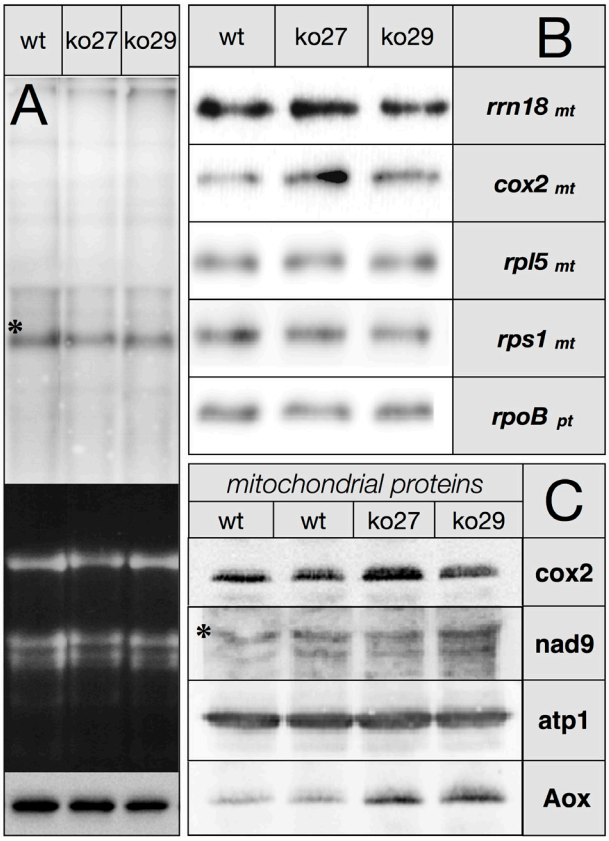


Figure 12: Transcript and protein expression in *PprpoTmp2* plants. In **A**, *atpB* Northern analysis of wt and knock-out lines (ko27, ko29) using a strand specific RNA probe (top) and RNA loading control (middle) as shown below are depicted. The bottom picture shows the result of the assertion of atpB steady-state protein levels by Western blot. When equal amounts of whole protein (20µg) were loaded no significant changes in atpB protein accumulation were observed. **B:** As for *atpB* additional Northern radio-labelling analysis did not show transcripts of unusual size for selected organellar genes, but an elevation of *cox2* was observed. In **C**, immuno-blot analysis shows accumulation of mitochondrial proteins in wild-type and mutant lines (20 µg protein per lane) with elevated levels in mutant lines for *A*lternative *ox*idase and slight increases for cox2. The asterisk denotes the specific nad9 signal.

(Dutilleul *et al.* 2003, Karpova *et al.* 2002, Noctor *et al.* 2004). Additionally, induction of Aox is commonly used as a marker for the mitochondrial retrograde response in plants beside its function in the plant mitochondrial electron transport chain. Investigation of Aox steady-state protein levels in mutant and wild-type protonema as shown in Figure 12 reveal an induction of Aox protein expression in *PprpoTmp2* plants possibly indicating a mitochondrial dysfunction.

3.1.4 Analysis of mitochondrial and plastid promoters in wild-type and *PprpoTmp2* plants

To study putative promoter specificities of the two phage-type RNAPs PpRpoTmp1 and PpRpoTmp2 in mitochondria and plastids and to identify plastid and mitochondrial promoters in *Physcomitrella patens*, organellar transcription initiation sites were experimentally determined using a modified 5'-RACE technique (Bensing *et al.* 1996). This method has been successfully applied to define primary transcript 5' ends in plastids (Miyagi *et al.* 1998, Swiatecka-Hagenbruch *et al.* 2008, Swiatecka-Hagenbruch *et al.* 2007) and plant mitochondria (Kühn *et al.* 2005). By this method tobacco acid pyrophosphatase (TAP)-treated RNA will yield PCR products for both primary and processed transcripts, whereas products resulting from primary transcript termini will be significantly reduced or absent in TAP untreated fractions. Thus primary transcripts can be identified by comparison of 5'-RACE products obtained from the TAP-treated and the untreated fraction of the same RNA (designated in this work as +T and –T lanes). While plant mitochondrial and/or plastidial transcription start sites have been experimentally defined in *Chlamydomonas*, *Euglena*, *Marchantia* and a variety of flowering plants (see 1.2.4 and 1.2.5), no such data exist for *Physcomitrella patens*.

As outlined in 1.2.5 plastidial genes exclusively transcribed from NEP promoters found to date are rare, including *accD*, *ycf2* in dicots and the *rpoB* operon in flowering plants. Information concerning differential usage of mitochondrial promoters by multiple mitochondrial phage-type polymerases found in eudicots and *Physcomitrella* is completely lacking. Considering the reported specificity of AtRpoTmp for selected promoters in plastids (Courtois *et al.* 2007, Swiatecka-Hagenbruch *et al.* 2008) and the common feature of various mitochondrial and plastidial genes to contain multiple initiation sites, promoter utilisation between *PprpoTmp2* and wild-type protonema cultures was compared. The rational here was to test if lack of transcription initiation from specific promoters can be observed in the mutant and therefore directly correlated to the lack of *PprpoTmp2* even though no reduced transcript levels were observed for mitochondrial or plastidial genes. Therefore analysis of plastidial and mitochondrial transcript 5’ ends was performed to detect putative changes in the promoter usage in *PprpoTmp2* plants. Figure 13, 14 and 15 provide graphical summaries of TSSs determined in this study using total cellular RNA prepared from liquid protonema cultures. Table 16 lists the mapped TSSs and their surrounding sequences. Promoters (P) and their corresponding TSSs are specified with the gene name and the position of the initiating nucleotide with respect to the start of the coding sequence (e.g., P*accD*-36).

3.1.4.1 Plastidial promoters

In both wild-type and *PprpoTmp2* plants two *accD* transcription start sites were identified that mapped to position -42 and -36 with respect to the *accD* translational start site (Fig. 13 and Tab. 16). The initiating nucleotide was found 8 (P*accD*-42) and 6 (P*accD*-36) nucleotides downstream from -10 regions typically observed in promoters recognised by the bacterial-type plastid polymerase (PEP), but no conserved -35 region is found in the upstream region of

P*accD*-42 and P*accD*-36. It is noteworthy here that the lack of a recognizable -35 box is also found in other plant PEP promoters and that a highly AT rich sequence is found directly upstream from both -10 boxes (see 1.2.5). When the TSS for *ycf2* was approached the only *bona fide* primary 5' end was mapping to the upstream region of *psbA*. In the plastid genome of *Physcomitrella psbA* is located upstream of the *ycf2* and *trnH* coding sequences. Therefore it must be inferred that in plastids of *Physcomitrella* protonema *ycf2* is dominantly cotranscribed with *psbA* and *trnH* from the *psbA* promoter P*psbA*-54. The mapping of the *rpoB* promoter revealed a major 5' end in very close proximity to the coding region. A signal observed in the +TAP reaction only was mapped to the region -28 upstream of the *rpoB* coding frame in both wild type and Pprpotmp2.

Table 16: Transcription initiation sites detected by 5'-RACE.

Promoter	Sequence	No. of clones (+TAP)	No. of clones (-TAP)
Pnad6-645	GT*GAAAAG*CGCTCATTTA**CATA**GCTAATTTATGA	14/16	2/8
Patp1-401	CTT*GAT*AATATTCAATTT**CGTA**GTAATTGCATGC	9/12	4/12
Pnad2-17	*GAAAA*TTTGATTAAAAAA**CATA**AAAAAAGAGGTTT	7/8	nd.
PtrnE-549	ATTACTACAAA*GAAAAAA***CATA**TATATATATGAA	12/12	nd.
Patp6-128	GTATTTCTTCATATATAA**TATA**TATATGAAAAAA	5/4/1/8/19	nd.
Pnad9as+414	*AAAA*CCTTCTACCAATAA**ATTA**AATAATTCCACA	11/12	6/12
Prrn26-153	C*GAAT*GATGCGATATGTA**AATA**GAAGATACGTAT	7/8	nd.
Prrn5-257	AGCTTAT*GAAT*GAATCAT**CATA**GATAATTCATTG	15/16	nd.
PtrnfM-215	AAATTATTCT*GAAA*CGCT**ATTA**AAAGTGATGAAC	8/8	nd.
PtrnfM-117	CGCCTAC*GAAT*GAATCAT**CATA**GAGACTCGCAAT	4/7/12	nd.
PtrnfM-36nc	GGAACTGCTGCGCAAAAAGAAATATTTTGCGGCG	15/18	nd.
PatpB-270	GTTTT*GAAT*GT**TAAACTAATTA**ATAAAAAGAAAA	5/8	nd
PatpB-144	*AGAAT*//TAT**TTATAT**AA**TATA**TCAATATAATAA	4/8	nd.
PatpB-214	TAAAAAAAAAG**TATAAT**AAACCTAGTTTATAATTA	4/8	nd.
PpsbA-54	ATACATTTTGTGTAA**TACTAT**AAATTAACAAGTT	16/18	nd.
PrpoB-28	TTCAATTTTATCAAT**TATAAA**AATATTTATATTTA'	14/16	5/8
PpsaA-161	ACCTTAAAGATGTGT**CATAAT**AAATTTGAATACCT	10/12	1/8
PptrnG-32	TCTTAAAAAAAACA**TATATAT**ATATGTATATACAT	7/8	nd.
PaccD-42	TTTTAAAAAGTAT**TAATTAT**TACTATATAAAAATT	7/16	2/8
PaccD-36	AAAGTATTAATTAT**TACTAT**ATAAAAATTAAATAA	9/16	3/8

Initiating nucleotides are underlined; putative promoter cores are written bold and the frequent (A)GAA(A) motif is highlighted. The number of clones sequenced for each promoter is given together with the frequency of the respective primary transcript 5' end as determined from *T*obacco *A*cid *P*hosphatase-treated (+TAP) and untreated (-TAP) RNA. Mitochondrial TSS are highlighted in grey. n.d. (not determined/ lack of -TAP signal).

Instead of a conserved NEP promoter motif the initiating nucleotide was found to be downstream from bacterial like -10 box typically found in PEP promoters. Moreover the initiating nucleotide is a thymidine as described for a variety of PEP promoters, whereas for the nuclear encoded phage-type RNA polymerases transcription initiation is highly conserved on A and G nucleotides. As described above for the accD promoters no conserved -35 region is observed but a AT-rich sequence is located directly upstream from the -10 region. When promoter mapping of *accD*, *ycf2*, and *rpoB* revealed no changes in promoter usage and observable upstream sequences were in favour of those genes being transcribed from bacterial like PEP promoters the plastid genome of *Physcomitrella* was screened for putative NEP like promoter sequences with the consensus motif YRTA. Only hits containing a box II GAA motif in close proximity upstream of the YRTA sequence (see 1.2.5) were subsequently chosen for promoter mapping. Such putative promoter sequences were found upstream of *trnG*, *atpB* and *psaA*. None of these motifs was found to drive transcription of the latter genes but P*atpB*-144. Instead bona fide bacterial like promoter consensus motifs are observable in

front of the TSS of *trnG*, *atpB* and *psaA*. In the case of the P*atpB*-144 the predicted NEP-like core motif YRTA is located at a position from -6 to -3 with respect to the adenosin (+1) at the TSS. Notably inspection of the P*atpB*-144 upstream region also reveals a sequence that exhibits similarity to bacterial like promoters (see Suppl. Fig.1).

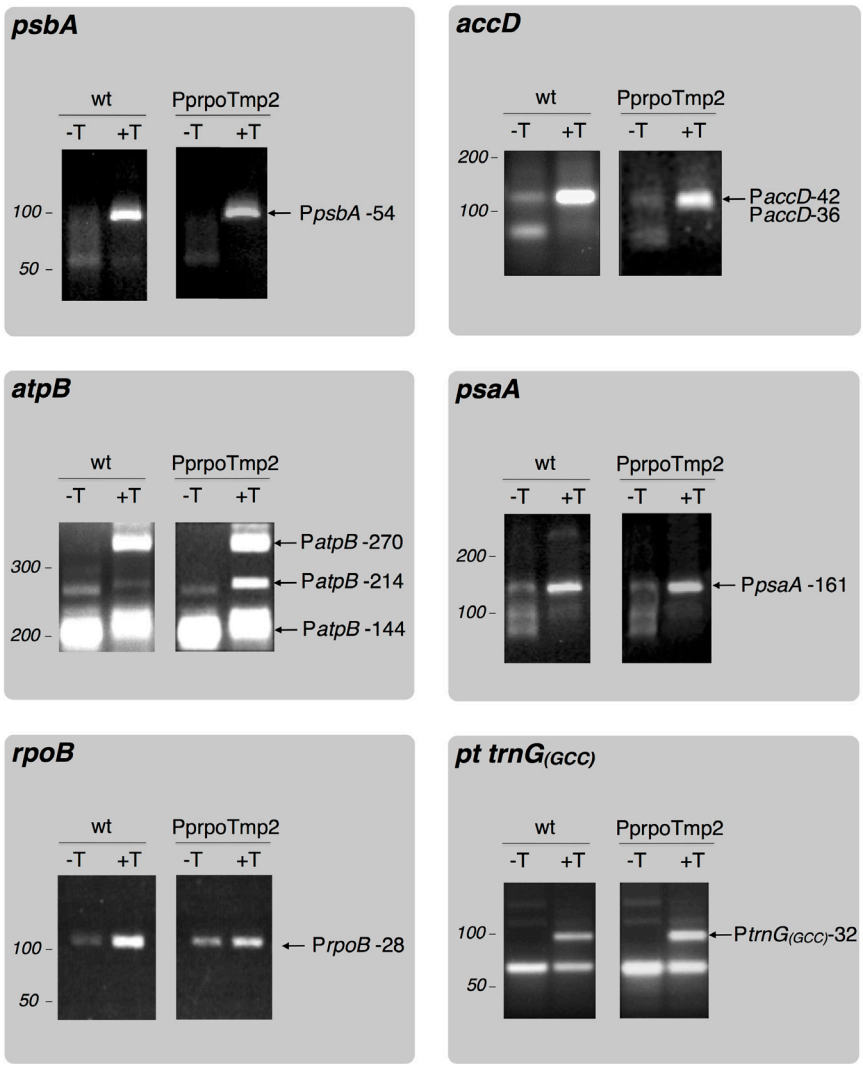


Figure 13: 5'-RACE analysis of the plastidial transcripts *accD*, *atpB*, *rpoB*, *trnG*(GCC), *psbA* and *psaA*. 5'-RACE products were separated on agarose gels; sizes are given in base pairs. Products corresponding to primary transcript 5' ends are indicated by arrows and labelled with the name of the corresponding promoter as listed in Table 16. Products resulting from primary transcript termini are recognizable since they are found to be significantly reduced or absent in TAP untreated fractions. All primary transcript 5' ends are identified in the wt and in PpRpoTmp2 plants as well.

However, promoter mapping in the Pprpotmp2 mutant did not show deviation in the usage of P*atpB*-144 and primer extension revealed no differences in the relative abundance of neither primary nor processed 5' ends (data not shown). Therefore, transcription from P*atpB*-144 is likely driven by PEP or PpRpoTmp1 in the mutant.

Since all plastid promoters active in the wild type were found to be active in *PprpoTmp2* as well, the data presented here do not provide evidence for a promoter specific function of PpRpoTmp2 in chloroplast transcription.

3.1.4.2 Mitochondrial promoters

Unlike for plastids no data were available for the specific function of multiple phage-type RNA polymerases in plant mitochondria as found in eudicots like *Arabidopsis* and the moss *P. patens*. As mentioned before a number of mitochondrial promoters have been mapped in flowering plants but no such data had been obtained for deeper branching land plants like *Physcomitrella*. To investigate the specific role of PpRpoTmp2 in mitochondrial transcription mitochondrial TSS identified in both mutant and wild-type plants were compared (see Tab. 16 and Fig. 14). *In silico* analysis of the mitochondrial genome of *P. patens* revealed a number of putative promoters matching the conserved CRTA core motif as identified in mitochondrial promoters of mono- and eudicotyledonous plants.

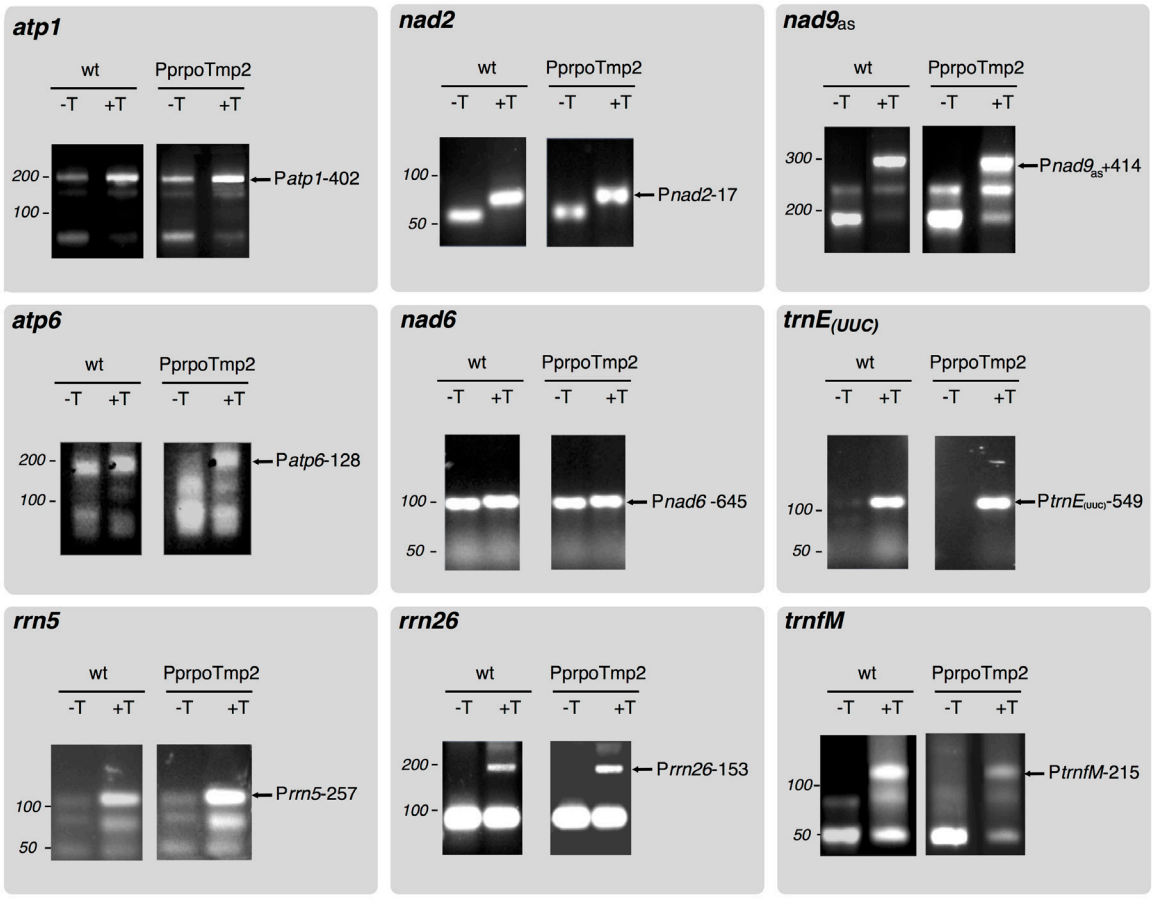


Figure 14: 5'-RACE analysis of the mitochondrial *atp1*, *nad2*, *nad9*, *atp6*, *nad6*, *trnE*(UUC), *rrn5*, *rrn26*, and *trnfM* transcripts. 5'-RACE products were separated on agarose gels; sizes are given in base pairs. Products corresponding to primary transcript 5' ends are indicated by arrows and labelled with the name of the corresponding promoter as listed in Table 16. Products resulting from primary transcript termini are recognizable since they are found to be significantly reduced or absent in TAP untreated fractions (-T). All primary transcript 5' ends are identified in both wild-type (wt) and *PprpoTmp2* plants.

Such putative promoter sequences were identified upstream of mitochondrial genes encoding *atpI, atp6, nad2, nad6, rrn5, rrn26, trnE,* and *trnfM,* as well as in antisense orientation downstream of the *nad9* stop codon. The predicted promoters of *atp1, nad2, nad6, rrn5, and trnE* were indeed found to precede the identified TSS. For *atp6, nad9as, rrn26,* and *trnfM*

TSS mapped to other loci diverging from the CRTA consensus. The TSS for the latter were found 2-4 nucleotides downstream from a WWTA motif also reported to precede mitochondrial transcriptional initiation sites in mono- and eudicotyledonous plants. Interestingly, inspection of nearly all mitochondrial promoters identified here revealed a GAAA motif in variable distance upstream to the core promoter sequences, reminiscent of the GAA box described for plastid type-Ib NEP promoters (see 1.2.5). Further examination unveiled the existence of this sequence motif in a variety of known mitochondrial promoters from mono- and eudicotyledonous plants as well.

The identification of an active promoter downstream of the *nad9* coding frame, driving transcription of a *nad9* antisense transcript was rather unsurprising since only recently synthesis of antisense RNA was demonstrated for the mitochondrial genes *atp9*, *nad4*, *nad5*, and *nad7* from *Arabidopsis* (Kühn 2006, Kühn *et al.* 2005). These presumably non-functional RNAs are thought to be the result of the complex structure of a frequently recombining mitochondrial genome, combined with its relaxed control of transcription (Holec *et al.* 2008b, Kühn 2006, Kühn *et al.* 2005). On the other hand, a fraction of these transcripts could be involved in the regulation of gene expression of mitochondrial genes, as was shown for a plastid antisense RNA that can protect the *Chlamydomonas atpB* transcript from exonuclease activity (Nishimura *et al.* 2004). All TSS identified in the wild type were also detected in the mutant, indicating that PpRpoTmp1 is the polymerase responsible for the transcription of this subset of mitochondrial genes or that PpRpoTmp1 is able to substitute for PpRpoTmp2 in the mutant.

Though all promoters were found to be active in *PprpoTmp2* plants, mapping of *trnfM* primary 5' ends revealed two additional transcriptional initiation sites only observable in the mutant (see Fig. 15). While in the wild type the primary 5' end was found to start at -215 with respect to the *trnfM* coding region, two additional TSS mapped to positions -117 and -36. The appearance of additional promoters in mutant plants is in concordance with the observed elevation of mitochondrial steady-state transcripts. Since all promoters identified in the wild type were also found to be active in *PprpoTmp2*, the data presented here do not provide evidence for a promoter specific function of PpRpoTmp2 in mitochondrial transcription.

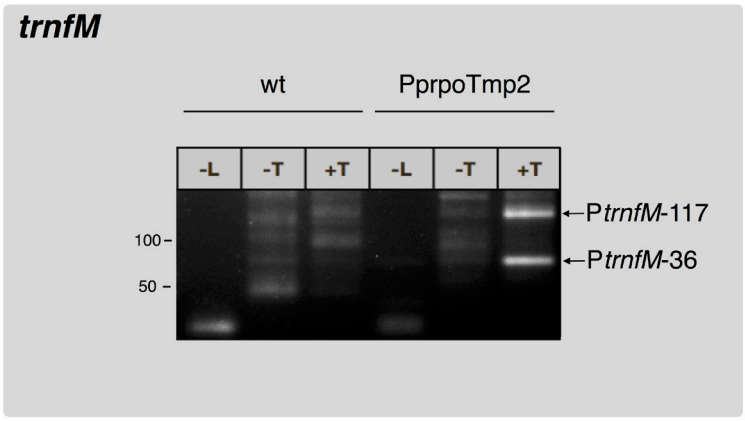


Figure 15: In *PprpoTmp2* plants two additional TSS are found for mitochondrial *trnfM*. Both primary transcript 5' ends are indicated by arrows and labelled with the name of their corresponding promoter (see Tab. 16). Lane -L corresponds to an additional control reaction lacking T4 RNA-ligase.

3.1.5 *In vitro* transcription studies of *Physcomitrella* RpoTmp1 and RpoTmp2

Only recently a major break-through was achieved by our group with the development of an *in vitro* transcription system for *Arabidopsis* phage-type RNA polymerases, allowing studies of the promoter specificities of the three *Arabidopsis* phage-type RNAPs and demonstrating that like the T7 RNAP, AtRpoTm and AtRpoTp can act as single-polypeptide enzymes *in vitro* and recognize numerous mitochondrial promoters and at least one plastidial promoter, while AtRpoTmp, despite having diverged from AtRpoTm more recently than AtRpoTp, is lacking this ability, with the exception of visualisable initiation at the mitochondrial promoter Patp6-916/913. To characterize putative differences in the promoter specificity and to the test for the ability of both *Physcomitrella* RpoT enzymes to faithfully recognise organellar promoters *in vitro*, PpRpoTmp1 and PpRpoTmp2 were heterologously expressed in *E. coli* using the pCold expression system (Takara Bio Inc.). In the *in vitro* transcription system of *Arabidopsis* phage-type RNA polymerases N-terminal thioredoxin-$(His)_6$ tagged proteins were used. To omit possible effects of the rather large thioredoxin tag on RNA polymerase activity, the expression vector pCold I was used. Subsequent purification made use of a N-terminal His_6-tagged version of both of the enzymes lacking the N-terminal transit peptides (Fig. 16). Removal of the His_6-tag from both RpoTmp1 and RpoTmp2 led to instability of both untagged RNA polymerases and therefore prompted His_6-tagged RpoTmp1 and RpoTmp2 to be employed for *in vitro* transcription experiments.

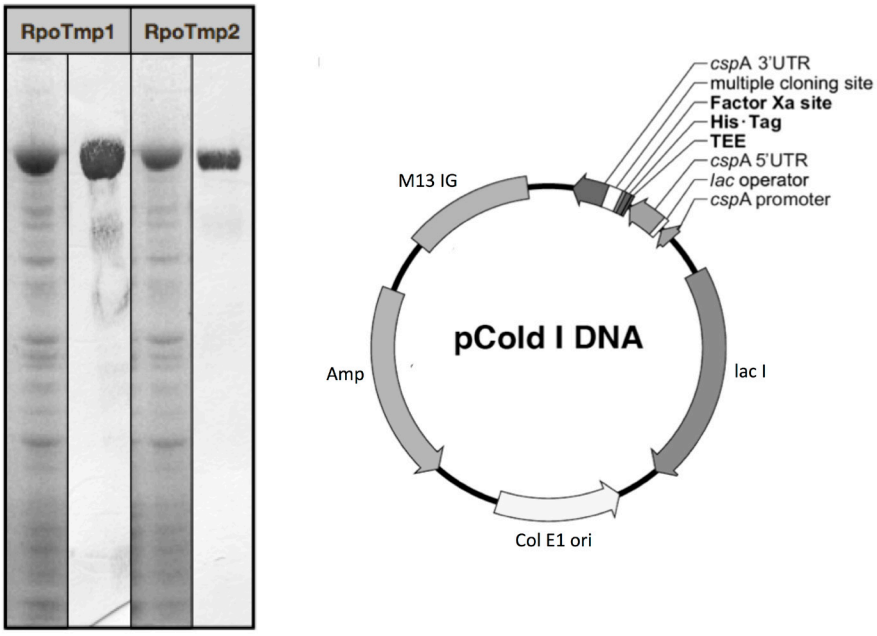


Figure 16: *Left* – RpoTmp1 and RpoTmp2 lacking the transit peptides and fused N-terminally to a hexahistidine tag were expressed in *E. coli*, purified by Ni^{2+}-NTA agarose, and analysed by SDS-PAGE followed by Coomassie-blue staining of the gels (lane 1 and 3), and by Western blotting and immuno-labelling of proteins with a polyhistidine antibody (lane 2 and 4). *Right* – corresponding cDNAs of both RpoTmp1 and RpoTmp2 were cloned into the multiple cloning site of expression vector pCold I; protein expression was induced by 0,5 mM IPTG and cold shock at 15°C.

Chapter 3.1.4 describes the mapping of transcription initiation sites in the plastid and mitochondrial genome of *Physcomitrella patens*. The knowledge, that these promoters are

recognised by the transcription machinery *in vivo* enabled the set up of an *in vitro* system. DNA templates for *in vitro* transcription were constructed by inserting promoter regions of plastid and mitochondrial DNA into pKL23 (Liere and Maliga 1999) upstream of the two bacterial terminator sequences *hisa* and *thra* (Barnes and Tuley 1983, Gardner 1982) present in pKL23 (Fig. 17). By providing a circular pKL23 derivative as a template in run-off experiments, transcription if initiated at the introduced promoters should be terminated at *hisa* and/or *thra* (compare Fig. 17), thereby producing RNAs of distinct lengths. Initially recombinant His_6-tagged PpRpoTmp1 and PpRpoTmp2 were assayed for faithful transcription initiation at the promoter sequences of *Physcomitrella* P*atp1*-402 (mitochondrial), plastidial *atpB* encompassing all three *in vivo* mapped promoters and P*psbA*-54, as well as the three mitochondrial promoters P*atp8*-228/226, P*atp6*-916/913 and P*rrn26*-893 as of *Arabidopsis* (Fig. 17).

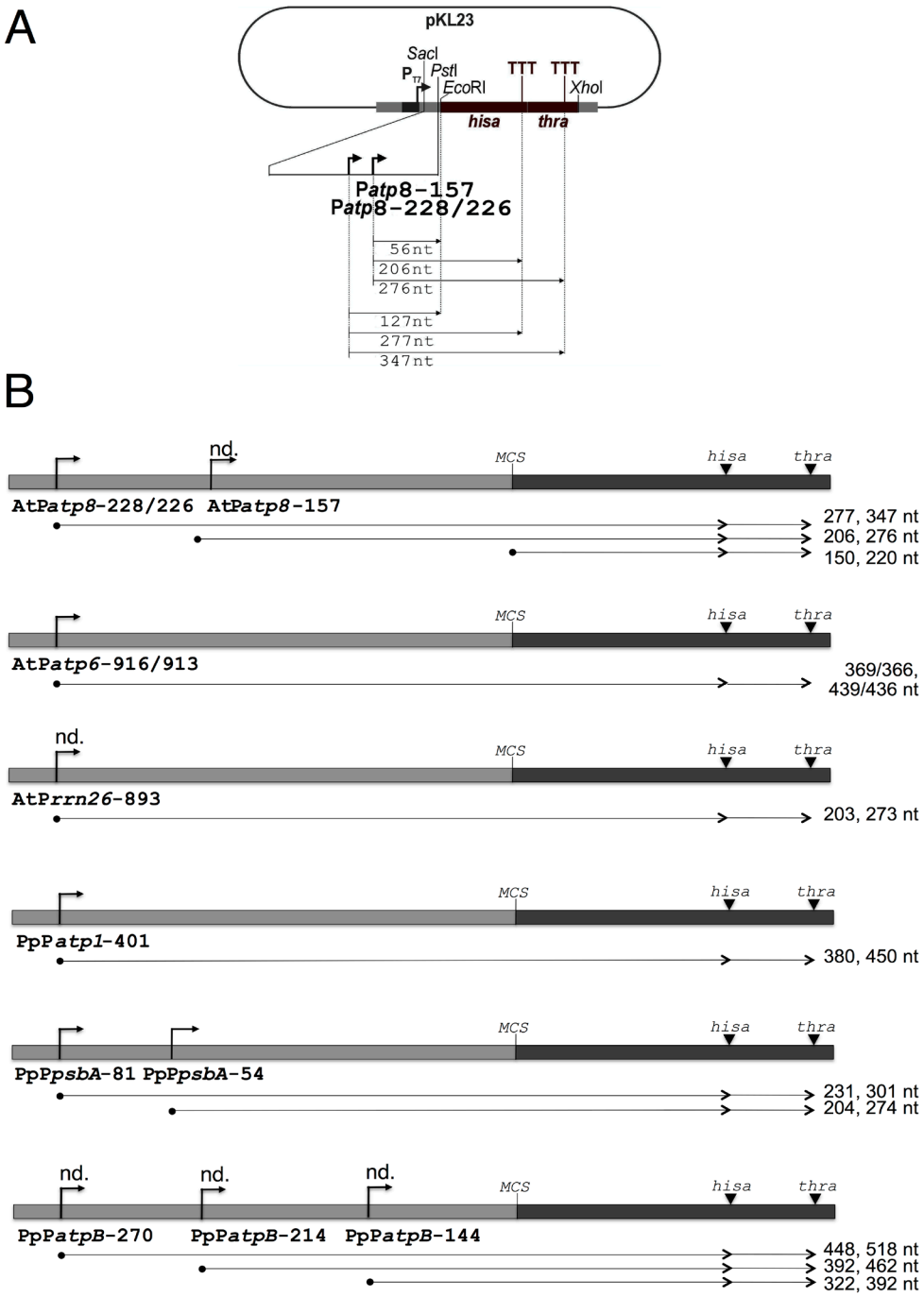


Figure 17: Generation of template plasmids for *in vitro* transcription. (A) pKL23-*atp8* was generated by inserting a 196-bp fragment of the *Arabidopsis* mtDNA containing promoters AtP*atp8*-228/226 and AtP*atp8*-157 (see B) into pKL23 via the SacI and PstI restriction sites as described before (Kühn 2006, Kühn *et al.* 2007). The positions of a T7 promoter (P_{T7}) represented by a dark grey bar and of the bacterial attenuators *hisa* and *thra* symbolised by red bars are indicated. Bent arrows mark start sites of transcription at AtP*atp8*-228/226, AtP*atp8*-157, and P_{T7}. **(B)** All other pKL23 template plasmids were generated as described for pKL-*atp8* above. Run-off products expected from initiation at AtP*atp8*-228/226, AtP*atp8*-157, AtP*atp6*-916/913, AtP*rrn26*-893, PpP*atp1*-401, PpP*psbA*-81, PpP*psbA*-54, PpP*atpB*-270, PpP*atpB*-214 and PpP*atpB*-144 and termination at *hisa* and *thra* are indicated by horizontal black arrows labelled with the respective RNA length. While black triangles mark the positions of terminator structures on the vector backbone (dark grey), mtDNA inserted into pKL23 is shown in pale grey.

The latter three *Arabidopsis* promoters where previously shown to be faithfully recognised by AtRpoTm and AtRpoTp *in vitro*, whereas AtRpoTmp initiation was only recognizable on autoradiographs when P*atp6*-916/913 was used in the assay (Kühn 2006, Kühn *et al.* 2007). To the contrary, *in vitro* transcription of pKL23-*atp6*-1-B confirmed initiation at P*atp6*-1-916/913 on supercoiled DNA by PpRpoTmp2 (Fig. 18), while no specific initiation was seen with PpRpoTmp1 (Fig. 18). The observation of a single signal in auto radiographs indicating efficient termination of transcription at the first terminator structure of pKL23 namely *hisa* is in line with data provided by the *Arabidopsis in vitro* transcription system (Kühn 2006).

While *Arabidopsis* polymerases AtRpoTm and AtRpoTp are reported to have a low preference for the promoter AtP*rrn26*-893 over random start sites, neither of the two *Physcomitrella* enzymes did recognize this promoter *in vitro* (Fig. 18).

As reported for the purely mitochondrial (RpoTm) and plastid (RpoTp) phage-type RNA polymerases from *Arabidopsis*, only P*atp8*-228/226 significantly supported specific transcription initiation by PpRpoTmp1 and PpRpoTmp2 while P*atp8*-157 residing on pKL23-*atp8* as well, was neither recognised by PpRpoTmp1 nor PpRpoTmp2 as a promoter (Fig. 18). Though PpRpoTmp1 was able to initiate on P*atp8*-228/226, non-specific RpoTmp1-driven transcription exceeded RpoTmp2 activity by far (Fig. 18).

Taken together data obtained for the mitochondrial promoter P*atp8*-228/226 of *Arabidopsis* are supportive of cross-species promoter recognition by both *Physcomitrella* phage-type RNA polymerases. Moreover, recognition of P*atp6*-1-916/913 by RpoTmp2, but not RpoTmp1 is indicative of differential properties of both enzymes. The general high portion of non-specific RpoTmp1-driven transcription is also observed for all other promoters tested and described hereafter and might therefore be considered as reminiscent of the characteristics described for the third phage-type RNA polymerases from *Arabidopsis*, RpoTmp (Kühn 2006, Kühn *et al.* 2007). The mitochondrial promoter P*atp1*-402 of *Physcomitrella* as mapped *in vivo* in this study, contains a well conserved CGTA core promoter motif typical for many plant mitochondrial promoters and was therefore chosen for initial studies. Transcription of pKL23-*atp1* by PpRpoTmp2 produced two major discrete RNA products of the expected lengths of approximately 380 and 450 nucleotides (Fig. 18). While the upper band resulted from transcription termination at *thra*, termination at *hisa* produced a much stronger signal indicating efficient termination at the *hisa* terminator structure. Moreover, PpRpoTmp1 was also able to initiate on P*atp1*-402, though non-specific transcription was clearly higher as for PpRpoTmp2. As a result of the lower ratio of specific to unspecific transcriptional initiation as seen for PpRpoTmp1, only the signal derived from termination at the first terminator sequence (*hisa*) is observable in the autoradiographs.

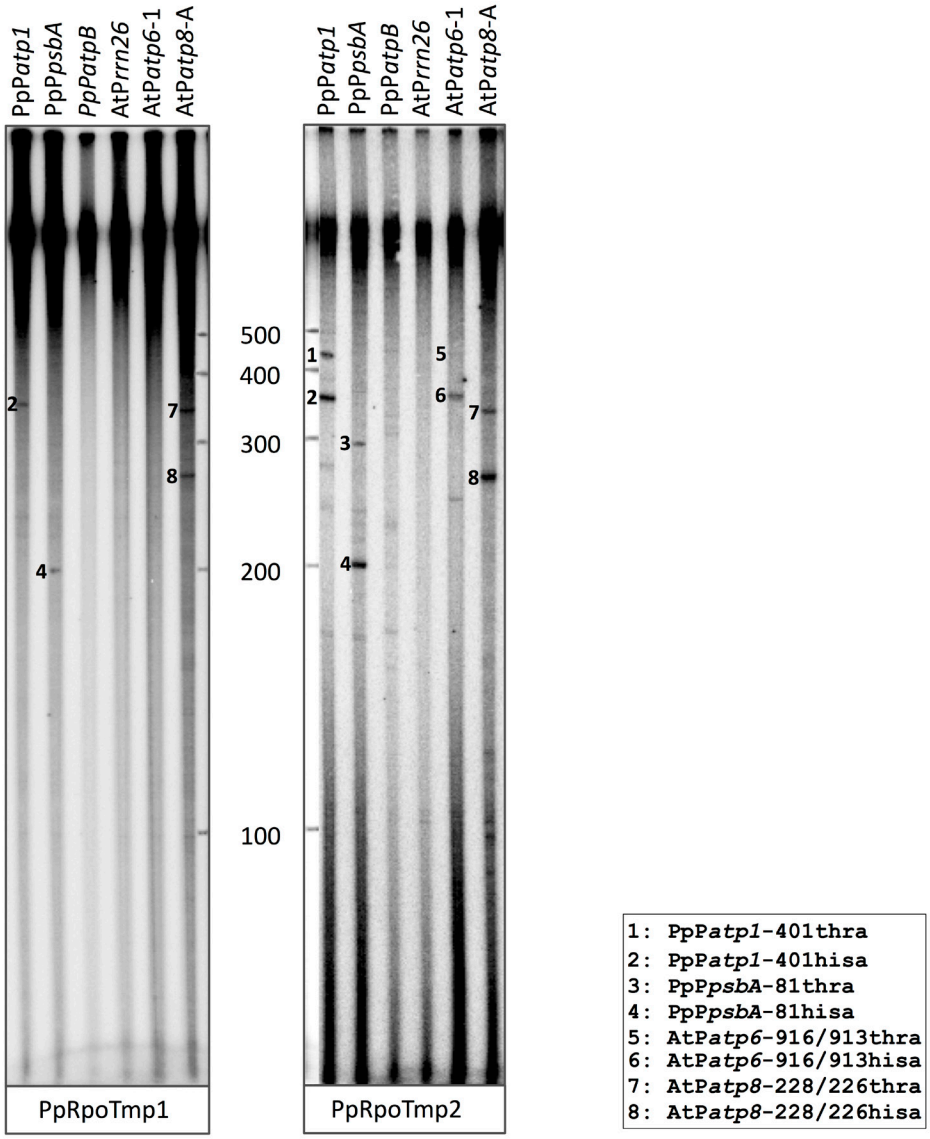


Figure 18: Run-off transcription from Physcomitrella and Arabidopsis upstream regions by PpRpoTmp1 and PpRpoTmp2. PpRpoTmp1 and RpoTmp2 were assayed for promoter-specific transcription on pKL23-*atp1*, pKL23-*psbA* and pKL23-*atpB* as of *Physcomitrella* and pKL23-*atp6*-1-B, pKL23-*atp8* and pKL23-*rrn26* from *Arabidopsis*. For *in vitro* transcription reactions with recombinant proteins RpoTmp1 or RpoTmp2 (400 fmol each) supercoiled (ccc) DNA was used and [^{32}P]-UTP-labelled RNA products were analysed after electrophoresis on a 5% sequencing gel alongside an RNA size standard; sizes are given in nucleotides. Transcripts made by RpoTmp1 were diluted 1:5 prior to PAGE due to RpoTmp1 showing a markedly higher activity than RpoTmp2. In concordance with all three *Arabidopsis* RpoT enzymes neither PpRpoTmp1 nor PpRpoTmp2 recognised promoters tested when plasmids were linearised (data not shown). Major discrete RNA products of expected length (see Fig. 16), are indicated by numbers 1 to 8 and listed on the right.

While the plastidial promoter of *psbA* P*psbA*-54 from *Physcomitrella* was included as a control for a typical PEP promoter containing both conserved -35 and -10 regions, the *atpB* promoter region was chosen since inspection of the upstream sequence of P*atpB*-144 shows similarity with both NEP and PEP promoter sequences (see Suppl. Fig. 1). The pKL23-*atpB* construct included all *atpB* promoters, none of which was used by either PpRpoTmp1 or PpRpoTmp2 in the *in vitro* transcriptional assay. pKL23-*psbA* surprisingly gave rise to a signal slightly larger than expected, if assuming transcriptional initiation at P*psbA*-54 and termination at the *hisa* or *thra* sequences. Moreover, structural elements on pKL23-*psbA* did allow for initiation of both *Physcomitrella* enzymes.

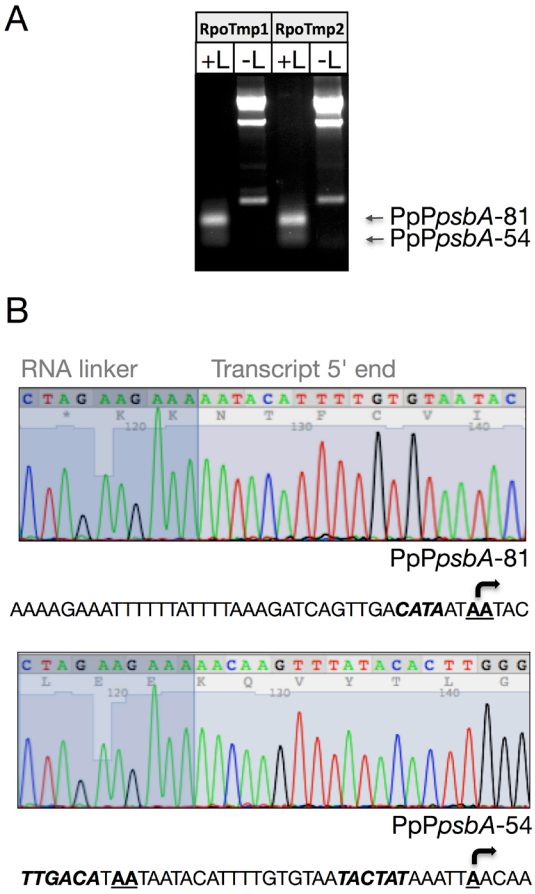


Figure 19: (**A**) 5’ end mapping was performed on RNAs synthesised by RpoTmp1 and RpoTmp2 from pKL23-*psbA*. RNA derived products are labelled with the corresponding promoter name, notably P*psbA*-81 is only mapped *in vitro*; specificity of RNA linker-ligated transcripts (lane +L) is also supported by comparison with products amplified from control samples (non-ligated transcripts, lane -L). In **B** the chromatograms display the sequence at the ligation site of a typical cloned 5’-RACE product; RNA linker and transcript portions of the sequence are indicated above and sequences preceding the initiation site are shown below. Initiating nucleotides are indicated by bold, underlined letters and bend arrows mark the initiation site specific for P*psbA*-81 or P*psbA*-54 as mapped from *in vitro* transcription reactions of both PpRpoTmp1 and PpRpoTmp2. Typical structural elements of organellar promoters are in italics.

To clarify if these signals were indeed specific for the promoter P*psbA*-54, 5’ termini of *in vitro* synthesised transcripts were identified as described in section 3.1.4 (Fig. 19, lane +L).

To furthermore distinguish RNA-derived PCR products from signals resulting from non-specific amplification of sequences from the carried-over DNA template, *in vitro* transcription products not ligated to the RNA oligonucleotide linker were subjected to RT-PCR as well (Fig. 19, lane -L). Electrophoresis of 5’-RACE products showed two bands a very minor lower and a major band, indicating that transcription initiated at two promoters (Fig. 19). While the method applied here is not quantitative, lack of a second signal in the autoradiographs (see Fig. 18) as well as the weak PCR signal (Fig. 19 A) suggested that initiation at the downstream promoter is very weak.

Sequencing of the cloned PCR products revealed that the transcripts resulting from the lower initiation site had a 5' end identical to that of transcripts initiated at P*psbA*-54 *in vivo*, allowing to attribute the *in vitro* synthesised RNAs to transcription initiation at a sequence clearly exhibiting all features of a *bona fide* PEP promoter (Fig. 19 B and Suppl.Fig.1). On the other hand, *in vitro* transcription and 5'-RACE experiments together provided evidence for the majority of the transcripts to be initiated specifically 81 nucleotides upstream of the *psbA* translational start site. Re-examination of the *psbA* sequence from P*psbA*-54 revealed the sequence motif TATA to be located 8 nucleotides upstream from the initiating nucleotide. Therefore, the TATA core element as part of the conserved -10 region of the supposed bacterial type promoter might drive initiation of transcription by both phage-type polymerases *in vitro* (Fig. 19 B and Suppl.Fig.1).

The major signal on the other hand corresponds to an initiating nucleotide that is preceded by a CATA core motif and a very AT-rich upstream region. Interestingly the core CATA motif of the *in vitro* promoter P*psbA*-81 is part of the -35 region of the putative PEP promoter region as mapped *in vivo* (Fig. 19B). Further attempts to identify primary transcripts corresponding to P*psbA*-81 *in vivo* by primer extension and the TAP based method were unsuccessful (data not shown). Since for both methods only RNA isolated from liquid protonemata cultures were applicable, the importance of the *in vitro* promoter P*psbA*-81 for plastid transcription under different growth conditions and at different developmental stages must remain an open question.

As shown in Figure 20, PpRpoTmp2 was able to recognize the *in vivo* mapped promoters P*rrn26*-153, P*atp6*-128, P*trnE*-549 and P*trnfM*-117 in the *in vitro* assay, while PpRpoTmp1 did not efficiently initiate transcription at those promoters. Together these data are supportive of PpRpoTmp1 and PpRpoTmp2 being functional RNA polymerases that both posses the inherent ability to recognize organellar promoters in a minimal *in vitro* transcription system without the aid of additional cofactors. Furthermore, it is interesting to note that the *Physcomitrella* enzymes were able to recognize mitochondrial promoters from *Arabidopsis in vitro*.

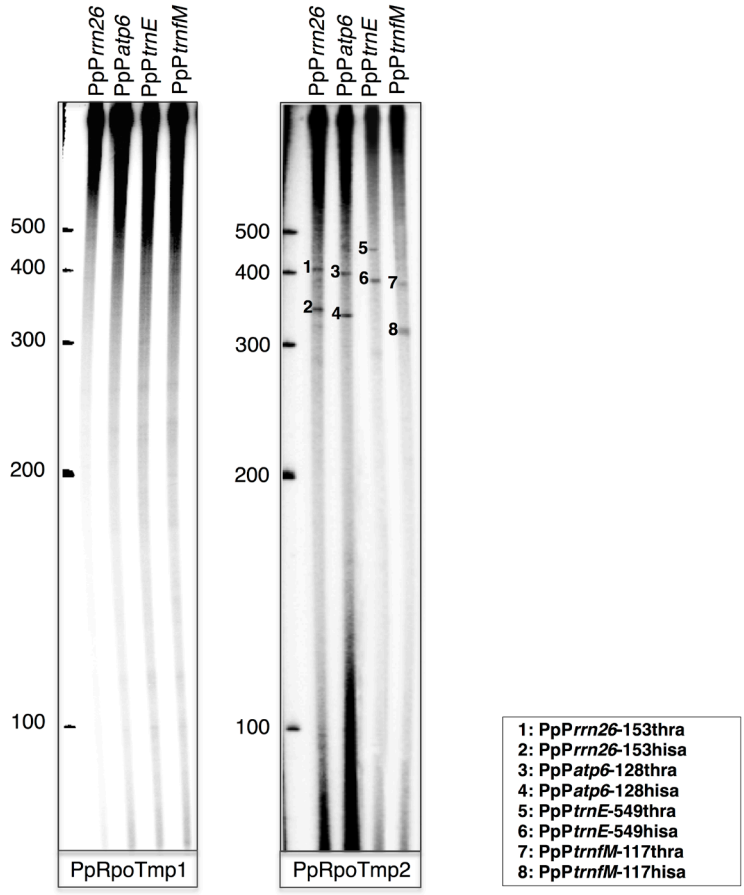


Figure 20: Run-off transcription from *Physcomitrella* upstream regions by PpRpoTmp1 and PpRpoTmp2. Mitochondrial DNA including the *in vivo* mapped promoters of *rrn26*, *atp6*, *trnE,* and *trnfM* were cloned into pKL23 as described before. Major discrete RNA products of expected length (see Suppl. Fig. 7), are indicated by numbers 1 to 8 and listed on the right.

3.1.6 Parenthesis - a third phage-type RNA polymerase in *Physcomitrella*

When the genome sequence of *Physcomitrella patens* was published (Rensing *et al.* 2008) *blast* analysis revealed a third genomic locus with strong similarity to plant phage-type RNA polymerases. Northern analysis revealed expression of all three *RpoT* open reading frames as shown in Figure 21. The full-length PpRpoT3 cDNA sequence (3496 bp) was obtained by aligning fragments produced by RT–PCR and RACE, and constitutes 235 bp of untranslated leader, the protein coding sequence of 3030 bp, and a 3'-untranslated sequence of 231 bp. Therefore, the predicted PpRpoT3 protein comprises 1010 amino acids and it shows 46–48% identity with the three *Arabidopsis* RpoTs.

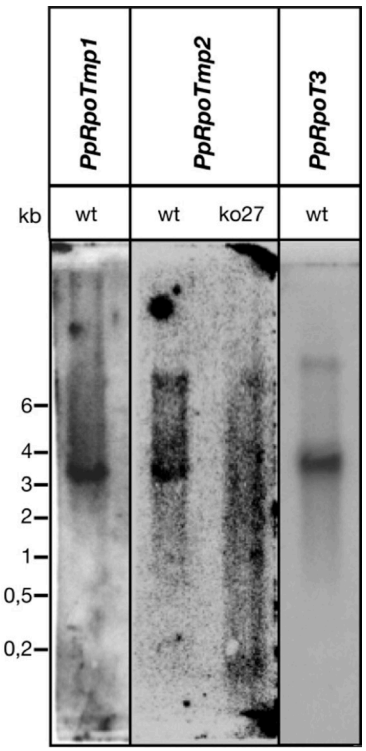


Figure 21: Northern analysis of all three *Physcomitrella* RpoT transcripts. Probes specific to the coding sequence of PpRpoTmp1, PpRpoTmp2 and PpRpoT3 were hybridised to 10 µg of filter-immobilised polyA-RNA isolated from wt and *PprpoTmp2* protonemal cultures. RNA size markers were run alongside samples; marker sizes are indicated on the left. Signals corresponding in size to expected mature transcripts are observed between three and four kb for all three *Physcomitrella* RpoT transcripts. As expected mature full length transcripts for *RpoTmp2* are lacking in the *PprpoTmp* RNA population.

While the two *Physcomitrella* RpoT polypeptides RpoTmp1 and RpoTmp2 share 59% identity, the scores for RpoT3 are significantly lower - 53% with RpoTmp1 and 51 % with RpoTmp2. Interestingly, the 46-48% identity scores of RpoT3 with the three *Arabidopsis* RpoTs are higher then those of RpoTmp1 (43-47%) and RpoTmp2 (43-46%). Figure 22 shows an alignment of the *Arabidopsis* and *Physcomitrella* RpoT polymerases demonstrating the high degree of similarity and the conservation of functionally essential residues and motifs in PpRpoT3. Thus, the derived amino acid sequence is consistent with PpRpoT3 encoding a functional RNA polymerase. To elucidate the gene structure of the third *Physcomitrella* RpoT gene, the cDNAs was aligned with the genomic sequence taken from the JGI genome browser (http://genome.jgi-psf.org/). The two published *RpoT* genes *RpoTmp1* (7,3 kb) and *RpoTmp2* (8.7 kb) contain 18 and 20 introns respectively (Richter *et al.* 2002). In *PpRpoT3* (7.3 kb) only 17 introns are found. Compared with the *RpoT* gene structure of flowering plants, *PpRpoT3* contains one additional intron in the 5' none coding part of the gene, intron 4 and 16 (as of *Arabidopsis* RpoTm) are missing in *PpRpoT3* (Suppl. Fig. 3).

The insertion sites of all other common introns are precisely conserved relative to the aligned amino acids sequences between *PpRpoT3*, the two other *Physcomitrella* genes and those of *Arabidopsis* (Suppl. Fig. 3). Furthermore, no conclusive evidence concerning the phylogenetic relationship of the three *Physcomitrella RpoT* genes can be inferred from the exon-intron structure. While the acquisition of intron 1 (as of *PpRpoTmp1*) is in favour of a

closer relationship of *PpRpoTmp1* and *PpRpoTmp2*, *PpRpoTmp1* and *PpRpoT3* both share the lack of the highly conserved intron 16 (as of *Arabidopsis RpoTm*).

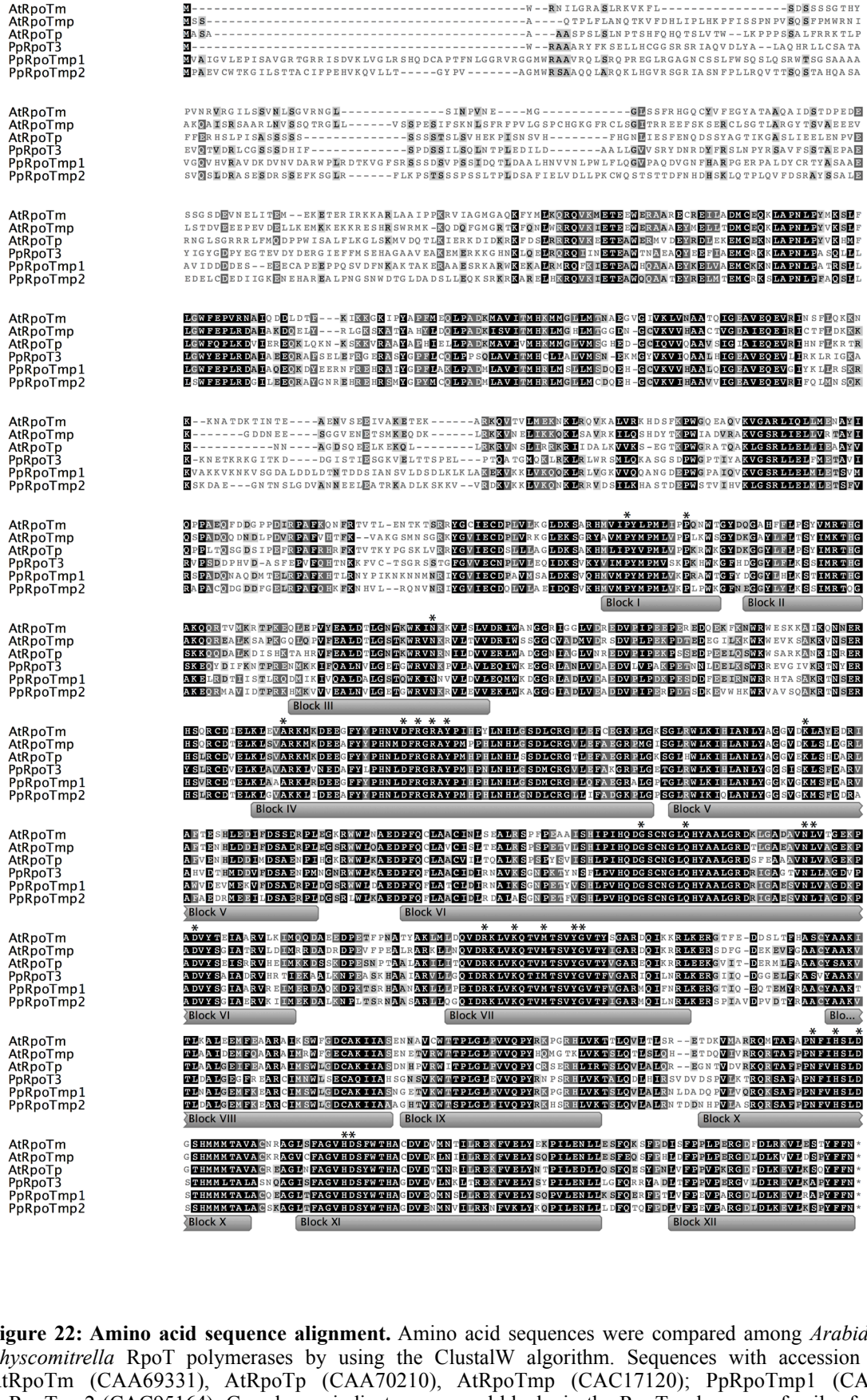


Figure 22: Amino acid sequence alignment. Amino acid sequences were compared among *Arabidopsis* and *Physcomitrella* RpoT polymerases by using the ClustalW algorithm. Sequences with accession numbers: AtRpoTm (CAA69331), AtRpoTp (CAA70210), AtRpoTmp (CAC17120); PpRpoTmp1 (CAC95163); PpRpoTmp2 (CAC95164). Grey boxes indicate conserved blocks in the RpoT polymerase family; functionally crucial residues (McAllister and Raskin 1993, Sousa *et al.* 1993) are indicated by asterisks.

3.1.6.1 Subcellular localisation of *Physcomitrella* RpoT3

While both *PpRpoT1* and *PpRpoT2* contain two in-frame AUG codons in their 5' ends, facilitating dual targeting properties due to alternative translational starts, for PpRpoT3 only one putative start codon was identified. Computer analysis of the putative transit peptide (Emanuelsson *et al.* 2000) produced high scores (0.8–0.9) for mitochondrial localisation. To experimentally investigate the subcellular localisation of the PpRpoT3-encoded polypeptide, the putative N-terminal transit peptide was fused in-frame to the GFP coding region. Furthermore, an additional construct was generated in such a way that it contained the whole native 5'UTR sequence, allowing for translational initiation at potential none AUG start-codons and to rule out GFP localisation due to forced translational initiation. After transfection of the PpRpoT3-GFP fusion constructs into *Arabidopsis* protoplasts transient expression was monitored by confocal laser scanning microscopy. Three plasmids encoding GFP only (Fig. 23B), a mitochondrial Cox4-GFP and a plastidial RecA-GFP (Fig. 23D) fusion protein were used for reference transformations of *Arabidopsis* protoplasts. As shown in Figure 23A protoplasts expressing PpRpoT3-GFP from both constructs displayed green fluorescence of small structures resembling the fluorescent mitochondria of protoplasts synthesizing Cox4-GFP (Fig. 23C), substantiating a mitochondrial localisation of PpRpoT3. The obtained data indicate, that the N-terminal part of PpRpoT3 has mitochondrial targeting properties and suggest that RpoT3 is exclusively targeted to mitochondria *in planta.* Therefore PpRpoT3 will be subsequently designated as PpRpoTm.

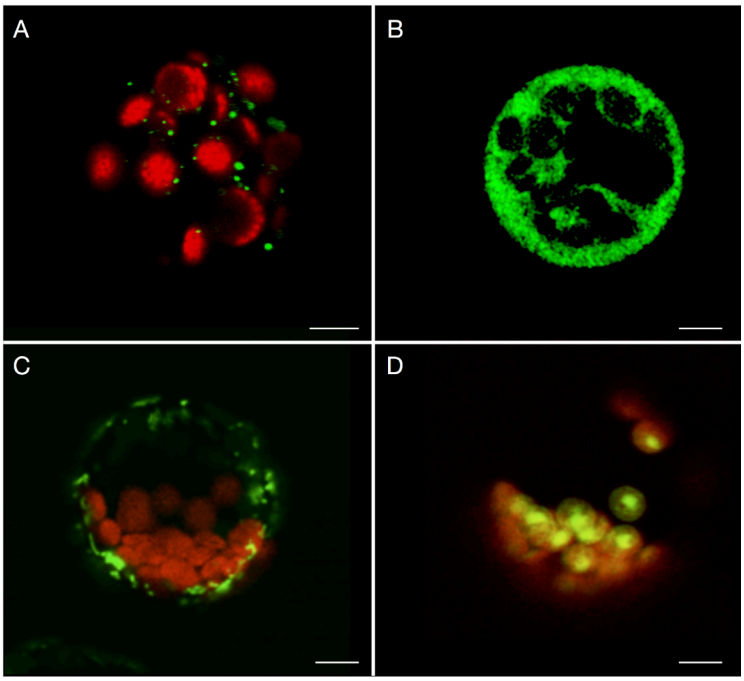


Figure 23: Transient expression of GFP fusion proteins in Arabidopsis thaliana protoplasts. The PpRpoT3 gene fragment encoding putative transit peptides was inserted into plasmid pOL-GFPS65C to generate a vector driving the expression of utrPpRpoT3 (A) showing mitochondrial GFP localisation. Control constructs code for GFP (B, pOL only), mitochondrial CoxIV-GFP (C) and plastidial RecA-GFP (D), respectively. Images were taken by confocal fluorescence microscopy. Scale bar: 10µm.

3.1.6.2 Expression of *RpoT* genes during the life-cycle of *Physcomitrella patens*

The expression of the *RpoT* genes in flowering plants is most likely influenced by various internal and external stimuli. While light regulation of all *Arabidopsis RpoT* genes was shown recently by T. Preuten (in preparation), Emanuel *et al.* demonstrated that the amount of *RpoT* transcripts varies between different tissues in flowering plants (Emanuel *et al.* 2006, Emanuel *et al.* 2004). Moreover, development-specific changes of transcript accumulation and promoter strength were detected for all three *Arabidopsis RpoT* genes (Baba *et al.* 2004, Emanuel *et al.* 2006). The self-fertile moss *P. patens* is a simple plant whose primary life stage consists of filamentous protonemal cell stage, with a succession of chloronema and caulonema cells and small leafy shoots (gametophores), all of which are haploid. On the apices of the gametophores both sex organs archegonia (female) and antheridia (male) are formed. After fertilisation the diploid mainly heterotrophic sporophyte grows from the zygote and is supported by the photoautotrophic gametophyte. Nothing is known about a potential regulation of plastid or mitochondrial transcription throughout these developmental stages. Quantitative real time RT-PCR analysis was facilitated to provide evidence for changes in the expression profile of all three *RpoT* genes throughout the *Physcomitrella* life-cycle. Since it is not known if normalisation can rely on standard transcripts usually used such as nuclear 18S rRNA or ubiquitin, when comparing transcript steady-state levels between different stages of the *Physcomitrella* life-cycle, *PpRpoT* transcript levels measured in gametophytes were used as the internal standard. While, changes for *PpRpoTmp2* appear less pronounced throughout the *Physcomitrella* life-cycle, as shown in Figure 24, expression of *PpRpoTmp1* and *PpRpoTm* genes is likely induced in sexual gametophytes and reduced at the protonema stage when compared with the expression of both genes in gametophytes. Since the RNA prepared from sexual gametophytes contains RNA pooled from late gametophytes, archegonia, antheridia and sporophytes it has to be assumed, that *PpRpoTmp1* and *PpRpoTm* gene expression is elevated at least during one of these stages.

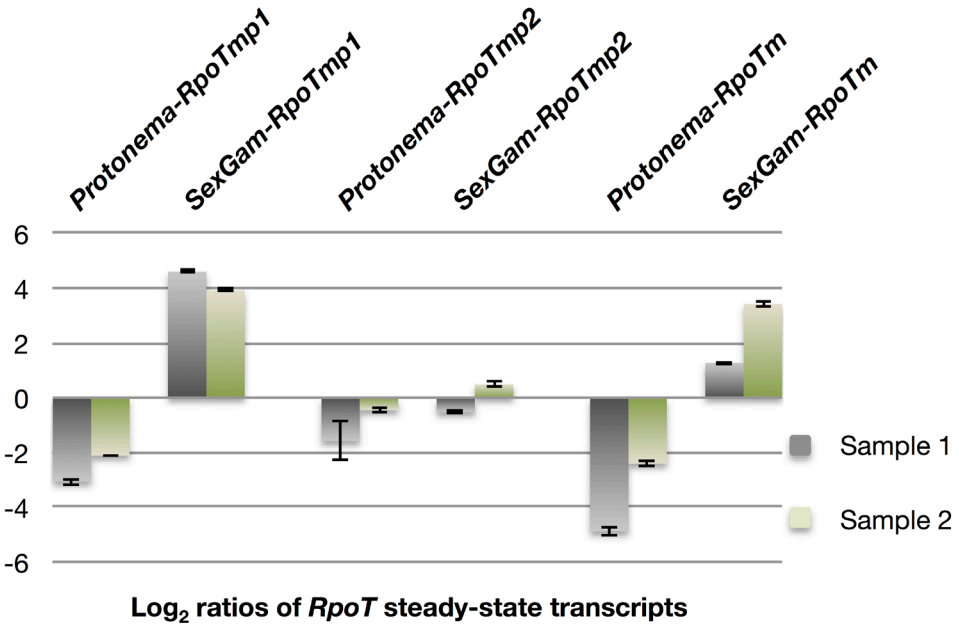


Figure 24: Expression of all *PpRpoT* genes in protonema and sexual gametophytes relative to the gametophytic phase. PCR was performed in triplicates and in two independent experiments (Sample 1 and 2).

3.1.6.3 Steady-state transcript levels of PpRpoTmp1 and PpRpoTm in *PprpoTmp2* plants

The *PprpoTmp2* mutant plants showed a general elevation of mitochondrial transcript accumulation and while none of the mitochondrial transcripts tested was reduced or lacking, instead additional promoters were found to be active. This suggests that the function of PpRpoTmp2 in mitochondrial transcription might be compensated by another mitochondrial RNA polymerase activity and overcompensation may account for the elevation of mitochondrial steady-state transcripts. To test this possibility, the transcript levels of *RpoTmp1* and *RpoTm* were investigated by quantitative real time RT-PCR from protonemal cultures. As apparent from Figure 25 the transcript accumulation of both *RpoTmp1* and *RpoTm* in the mutant was higher when compared to the wild type. Moreover the elevation of *RpoTmp1* transcripts (5- to 6-times) was much more pronounced compared to the elevation of *PpRpoTm* transcripts (2-times). Both findings together may suggest the existence of a regulatory mechanism to compensate for the loss of RpoTmp2 function in the mutant and possibly are the direct cause for the observable elevation of mitochondrial transcripts.

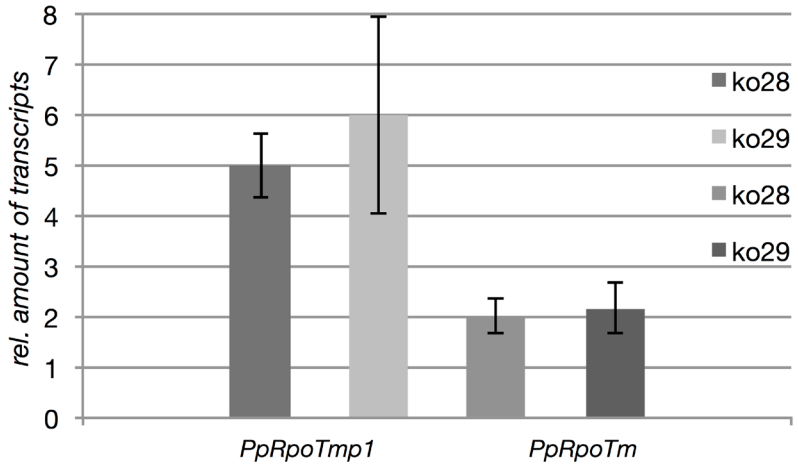


Figure 25: Quantitative real-time RT-PCR analysis of *PpRpoT* gene expressions. Relative expression of PpRpoTmp1 and PpRpoTm genes in *PprpoTmp2* lines 28 and 29 compared to the wild type. Cultures were grown for 10 days under standard conditions as for macro-array analysis. PCR was performed in triplicates and in two independent experiments with 18S rRNA as internal control. Averaged results are shown in the graph. Bars represent SD.

3.2 Analysis of deviating functions of organellar phage-type RNA polymerases in *Arabidopsis thaliana*

The *AtRpoTm* and *AtRpoTmp* genes in *Arabidopsis* have been reported to display overlapping expression patterns in different tissues and at different developmental stages (Emanuel *et al.* 2006) and were therefore proposed to recognize different types of mitochondrial promoters. In a study of an *Arabidopsis* line lacking AtRpoTmp (Baba *et al.* 2004) alternatively both AtRpoTm and AtRpoTp have been suggested to be important at a later developmental stage, while AtRpoTmp was proposed to be the key RNA polymerase transcribing organellar genes during early seedling development. However, plants lacking AtRpoTmp displayed no changes in the accumulation of several mitochondrial transcripts tested (Baba *et al.* 2004). Moreover, AtRpoTm, but not AtRpoTmp, has been shown to specifically recognize mitochondrial

promoters *in vitro* (Kühn *et al.* 2007). Therefore, the transcriptional role of AtRpoTmp in mitochondria was still unclear. In plastids, however, AtRpoTmp and AtRpoTp were suggested to have partially redundant roles since insertion lines with both enzymes disrupted were seedling lethal (Hricova *et al.* 2006) and plants lacking either AtRpoTmp or AtRpoTp showed similar phenotypical alterations when compared to the wild type. Together with the data presented herein for the two *Physcomitrella* enzymes distinct functions of multiple phage-type RNA polymerases in mitochondria still had to be assigned. This question was addressed in a joint afford by the group of Prof.Whelan and our group.

3.2.1 AtRpoTm disruption is lethal

Neither in our group nor in the laboratory of Prof. Whelan plants homozygous for a T-DNA insertion in *AtRpoTm* from two independent T-DNA lines, SALK_005875 and GABI_350F01 could be isolated. Transmission of *AtrpoTm* alleles by both male and female gametes were found to be reduced by two-thirds in reciprocally backcrossed plants in the laboratory of Prof. Whelan, indicating decreased *AtrpoTm* gamete fitness (Kühn *et al.* 2009).

Taken together, the observations made in both laboratories suggest that disruption of AtRpoTm is likely lethal, indicating an essential function of AtRpoTm in mitochondria. In contrast homozygosity of T-DNA insertions (T-DNA insertion lines GABI_286E07 (*rpoTmp-1*), SALK_132842 (*rpoTmp-2*), and SALK_086115 (*rpoTmp-3*)) in the *AtRpoTmp* locus was not lethal, but produced a phenotype characterised by a developmental delay and wrinkly rosette leaves (Baba *et al.* 2004, Courtois *et al.* 2007).

3.2.2 Mitochondrial transcript levels not transcript sizes differ between *AtrpoTmp* and Wild-Type Plants

As described before the rosette leaf morphology displayed by *AtrpoTmp* plants is reminiscent of the leaf phenotype described for mutants impaired in mitochondrial function (de Longevialle *et al.* 2007, Reichheld *et al.* 2007, Van Aken *et al.* 2007) and similar aberrations in leaflet morphology were observed for *PprpoTmp2* plants (see 3.1.2). Therefore, the accumulation of mitochondrial transcripts in *AtrpoTmp* seedlings and rosette leaves was analysed by quantitative RT- PCR (qRT-PCR) in the group of Prof. Whelan.

The specific qRT-PCR system, initially developed by the group of Prof. Small at the Arc Center of Excellence to screen for splice defects, covers all annotated protein coding and rRNA genes of the *Arabidopsis* mitochondrial genome and is able to analyse the independent transcription of exons of the trans-spliced *nad1*, *nad2*, and *nad5* mRNAs.

Strong changes in the abundances of numerous mitochondrial transcripts in three mutant lines were found consistently for *rpoTmp-1*, *rpoTmp-2*, and *rpoTmp-3* when compared to the wild type (Fig. 26). In 7-d-old seedlings, *matR*, *nad1e4*, *nad2e1_e2*, *nad6*, *cox1*, and *rps4* transcripts were reduced four-fold in all three mutant lines, while lesser reductions in transcript levels were observed for *nad9* and *ccmC*. In the light of the elevation of several mitochondrial transcripts in *Physcomitrella rpoTmp2* plants it was intriguing that numerous other transcripts in *AtrpoTmp* plants were found to be more abundant in mutant lines, with the most elevated (*ccmFc*) being increased threefold. Taken together the differences in transcript abundances between mutant and wild-type plants as found by Dr. Kühn indicate this molecular defect to be due to the lack of a functional AtRpoTmp enzyme.

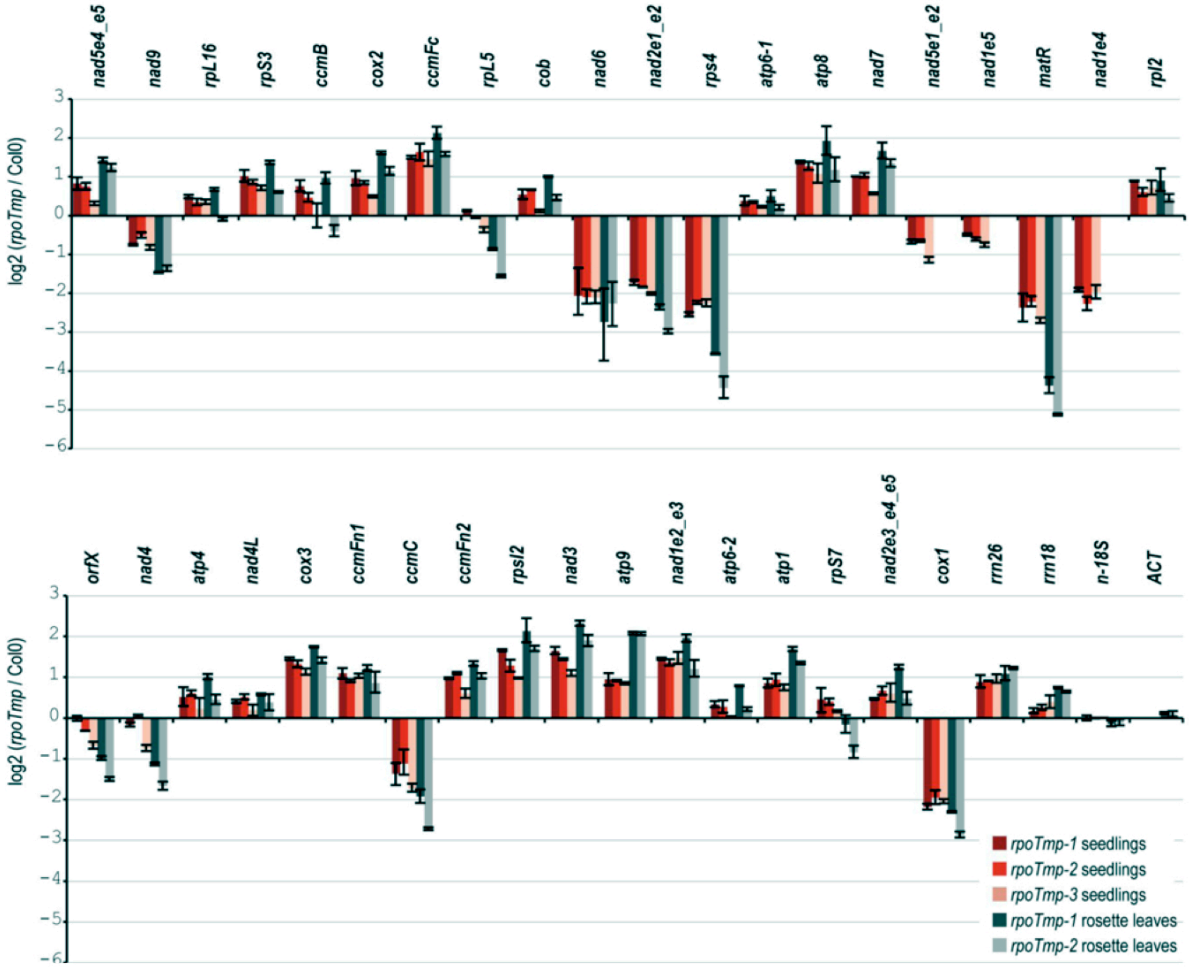


Figure 26: Mitochondrial Steady-state Transcript Levels in rpoTmp Mutants. Figure taken from Kühn et al. 2009. Log2 ratio of transcript levels in mutants compared with levels in wild-type (Columbia-0 [Col-0]) plants at the seedling and at the rosette stage. Three technical replicates are averaged per genotype; standard errors are indicated. Data normalised with nuclear 18S rRNA gene (seedlings) or the nuclear 18S rRNA and ACT genes (rosette leaves).

3.2.2.1 Validation of qRT-PCR data by strand specific Northern blot analysis

Since Baba and colleagues reported mitochondrial transcript abundances not to be affected in *AtrpoTmp* (Baba *et al.* 2004), the qRT-PCR results were validated through RNA gel blots. mRNA levels were confirmed to be reduced using cRNA probes complementary to transcripts *nad2*, *nad6*, *nad1* and *cox1*. Enhanced accumulation of the *ccmFc* mRNAs in rpoTmp mutants was consistent with the data provided by qRT-PCR (Fig. 27). While transcript levels were clearly shown to differ between *AtrpoTmp* and wild-type plants no transcripts of unusual size were seen to accumulate for the genes examined by RNA gel blot analysis. Since antisense transcripts appear to be a prevalent feature of metazoan and plant organellar transcriptoms (Burzio *et al.* 2009, Falkenberg *et al.* 2007, Georg *et al.* 2010, Holec *et al.* 2008b, Kühn 2006, Nishimura *et al.* 2004, Torres *et al.* 2009), deviations in the level of reduction between qRT-PCR and Northern results are likely due to the fact that qRT-PCR can not distinguish sense from antisense transcripts. Together strand specific DNA blot analysis confirmed changes indicated by qRT-PCR and did not provide evidence for accumulation of transcripts with unusual size.

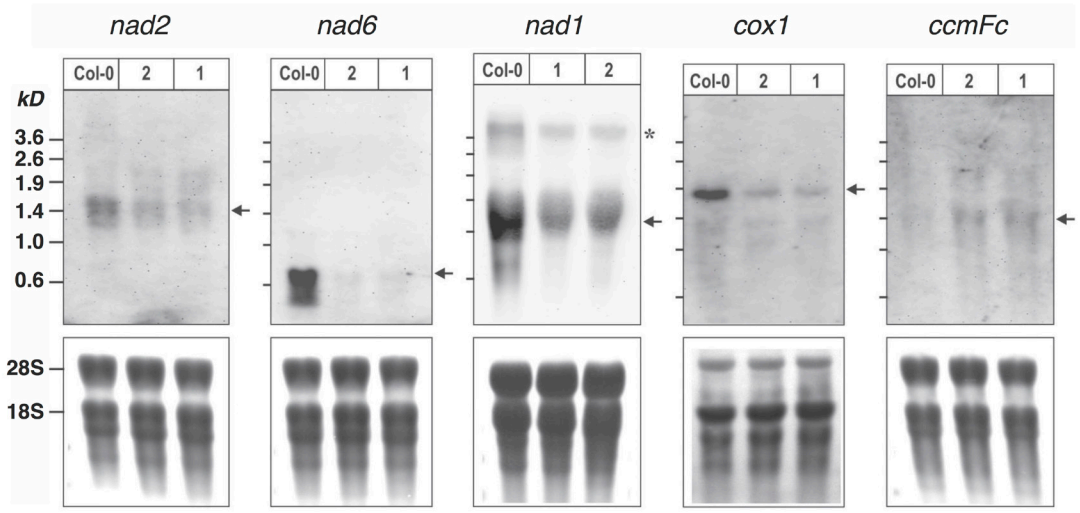


Figure 27: RNA gel blot hybridization of mitochondrial transcripts in rpoTmp. Filter-immobilised total RNA isolated from rpoTmp-1 (lane 1), rpoTmp-2 (2), and wild-type (Col-0) seedlings (top panels) were hybridised with complementary RNA probes specifically targeting the coding sequence of *nad2* (exon 1), *nad6*, *nad1*, *cox1*, *ccmFc* (exon1). RNA size markers run alongside samples are indicated on the left of each blot. Signals of the expected mature transcripts (Forner *et al.* 2007) are indicated by arrows to the right of each panel. As a loading control filters were stained with methylene blue prior hybridisation (bottom panels).

3.2.2.2 Identical promoters are utilised in *AtrpoTmp* and Wild-Type Plants

As observed deviations of mitochondrial transcript levels may result from changes in the rate of transcription but also could have been a consequence of differing rates of transcript degradation or turnover, mitochondrial run-on transcription assays were performed by Dr. Kühn to isolate the molecular defect (see Suppl. Fig. 4). While transcription of *nad6*, *nad2*, *rps4*, and *cox1* was significantly reduced in the mutant and rps3 did not show differences, run-on transcription of *rrn26*, *ccmFc*, and *atp6-1* in the mutant was significantly higher than in the wild type. The run-on transcription assays performed on the same set of plants as the qRT-PCR assays, revealed transcriptional changes in rpoTmp to coincide with the measured differences in transcript accumulation. Although reduction in transcription of *rps4* and *nad2e1_e2* in the mutant was not as severe as the measured decrease in transcript accumulation, indicating a contribution of altered transcript stabilities to changed transcript abundances in rpoTmp, transcriptional changes clearly account for the changes in transcript abundances in the mutant. Moreover lower levels of a specific set of mitochondrial transcripts likely due to reduced transcription were correlated with decreased abundances of the respiratory chain complexes I and IV (Kühn *et al.* 2009).

Considering the reported specificity of RpoTmp for selected promoters in plastids (Courtois *et al.* 2007, Swiatecka-Hagenbruch *et al.* 2008) and the aforementioned reduction of transcription, promoter utilisation between rpoTmp and wild-type seedlings was compared to test, if the reduced transcript levels observed for several genes in mutants were due to the lack

of transcription initiation from specific promoters, which in turn could be described as AtRpoTmp specific promoters. Apart from *cox1* (Kühn *et al.* 2005), no primary transcript 5' end or promoter of the genes displaying reduced transcript abundance and transcriptional performance in rpoTmp had been characterised. Therefore, a comprehensive examination of mitochondrial TSSs for genes displaying altered transcript accumulation and transcription in mutants was undertaken. As described before promoters in plant organelles can be identified by TSS mapping employing a 5'RACE technique (see 3.1.4). Table 17 lists the mapped TSSs and processing sites and their surrounding sequences. Promoters (P) and their corresponding TSSs are specified with the gene name and the position of the initiating nucleotide with respect to the start of the coding sequence (e.g., P*cox1*-355). If transcript 5' end analysis was not supportive of a functioning as TSS and the upstream sequence was devoid of elements frequently associated with mitochondrial promoters those sites are listed as processing sites (e.g. *cox1*-241). Nearly all of these termini have previously been identified as processing sites resulting from endonucleolytic 5'-terminal processing (Forner *et al.* 2007, Kühn *et al.* 2005). Eight 5' termini mapped to a promoter-like motif such as P*nad2e1*-114, but 5'-RACE analysis did not distinguish a primary 5' end (see Suppl. Fig. 5). These sites are still listed but as putative promoters ([e]) since it is known that some primary transcript ends are indistinguishable from processed ends by the 5'-RACE technique applied in this study (Forner *et al.* 2007, Kühn *et al.* 2005). These rare cases can be observed due to the *in vivo* instability of the 5' triphosphates and when primary transcripts are not processed to their mature forms at their 5' termini by exo- or endonucleolytic activities.

Table 17: Mitochondrial transcrition initiation and 5' processing sites in wild-type and *AtrpoTmp* plants.

			Col-0		*rpoTmp*	
Gene	5′ End	Sequence	+TAP	−TAP	+TAP	−TAP
cox1	P*cox1*-88	CTCTCAAGAACT**CGTA**GACTATGGA	2/8	0/7	n.a.	n.a.
cox1	*cox1*-241[a,b]	GAATCCAATGGTAGGCCTGGGCTCT	n.a.	n.a.	n.a.	n.a.
cox1	P*cox1*-355[a]	AATTTATTCA**ATTA**TATaATAATAA	n.a.	n.a.	6/8[#]	n.a.
nad1	P*nad1e4*-275	TACAAAAACAT**TCTA**GGaAAGGGCT	6/10	1/8	2/2	n.a.
nad1	P*nad1e4*-775	CGACTGCTA**TCTA**AGAGAAgATAAG	4, 3, 2/13	7, 0, 7/15	8/8	n.a.
nad1	P*nad1e4*-838	GGCAATTTAT**CGTA**TAGaAGAATAG	3, 6/12	0, 0/13	14, 0/14	n.a.
nad1	*nad1e4*-892	GACCAAGCAAATCTCCTCTTCTAGT	–	2/2	–	–
matR	P*matR*∼3*[,e]	CAATATGAAGAAAGAAaGAA*GGG*TT				
matR	P*matR*∼58*[,e]	GCCTTTG**ATTA**TAAAAAaGGGGAGC				
nad1	P*nad1e5*-427	ACTA**ATTA**TATATATATaATAAGGT	5/7	0/10	2/8	n.a.
nad1	P*nad1e5*-927	CAAATATTGTAAATCAAgAAGGTGG	5/18	0/10	2/13	n.a.
ccmC	P*ccmC*-45	CTTTGATTG**CGTATTA**TaGATCCAT	4/16	0/15	1/14	n.a.
ccmC	P*ccmC*-487[c,e]	TATCGATCGG**CGTA**AGaAAAgATTC	1, 2, 1/15	0, 2, 1/12	1, 1/8	n.a.
ccmC	P*ccmC*-1159	GATTCGTCAT**GGTA**GAGaAGAAAAA	6/6	–	8/8	–
ccmC	*ccmC*-1634	CCTCGAATTCTGTAAGCTAGCTTGG	1/1	n.a.	9/9	n.a.
ccmC	P*ccmC*-1677	CTACCAAACGTGCTAATAAAAGAGC	8/9	0/7	–	–
ccmC	P*ccmC*-1817	AAGTAATTC**ATTA**TAAgATA**AGTa**A	1, 3, 9/13	0, 0, 4/8	2, 2/9	n.a.
ccmC	P*ccmC*-1834[e]	ACACTACTATATAAGATAAGTAATT	2, 0/2	1,4/6	n.a.	n.a.
rps4	*rps4*+2[b]	GAGTCCCAGGGACGCA*At*GTGGCTG	13/13	n.d	8/8[#]	n.a.
rps4	P*rps4*-477	AGAAAGCTCA**AGTA**TGTaAACAACC	6/7	n.a.	8/8	n.a.
rps4	P*rps4*-1509	AGATAGGTCA**GGTA**TATaTAGCGAT	4/6	0/8	3/3[#]	–
nad2	P*nad2e1*-114[b,e]	CACTATC**ATTA**t**ATTA**TATaTTCAT	6,10, 4/32	5, 3, 2/18	3, 1/7	n.a.
nad2	P*nad2e1*-299	CAAGATATTG**CGTA**TaAATATATAT	2/28	n.a.	1/8	n.a.
nad2	P*nad2e1*-413	GAACAAAAGATGTAATGaAAAgATA	2, 3/10	n.a.	4, 3/8	n.a.
nad6	P*nad6*-88	GAAGAAAATGAAATAGGAaCAACCG	2,10/16	0,1/7	7/15	n.a.
nad6	P*nad6*-143	ATTGAGATTC**CGTAAGT**aACTCAGT	3/8	n.a.	9/16	n.a.
nad6	*nad6*-179[b]	GGGTACAAGATCGAAAAgAATGCAT	14/17	12/14	3/7	n.a.
nad6	P*nad6*-333	GATG**ATTA**TGTTGAAAAaGATCGAA	–	–	8/8	0/7
nad6	P*nad6*-401	CGGGTAC**AGTAGTAGGTa**AACTTGA	18/19	2/8	8/8	–
ccmFc	*ccmFc*-124[b]	AAGTCCCTCCTTCCGCTCCTGGTGT	n.a.	n.a.	n.a.	n.a.
ccmFc	P*ccmFc*-379[e]	TTGGATTAT**CGTA**GAATAAGAGGAG	6/7	6/7	n.a.	n.a.
ccmFc	P*ccmFc*-743	TTTATTATTCT**ATTA**TAaTAGTCTA	–	–	8/8	–
nad9	*nad9*-211/210[c]	GGCGATCCACGAATTTGATCATTCG	1, 2/8	n.a.	n.a.	n.a.
nad9	P*nad9*-1241[e]	TACGAAATAT**AGTA**AATATCGTAAA	3/3	n.a.	n.a.	n.a.
nad9	P*nad9*-1371	TTCTCGTAATGTATTTAACCATAAA	2/8	n.a.	–	–
nad9	*nad9*-1400	CAAACTAAATGCTATAAGCATAGAA	1/2	n.a.	n.a.	n.a.
nad9	P*nad9*∼1730[e]	GATTGATAAA**GGTA**ATAGATATGAT	n.a.	n.a.	n.a.	n.a.

Underlined: TSS and processing sites as mapped in the wild type (Col-0); lowercase positions indicate sites mapped for rpoTmp. ATTA, AGTA, CATA, CGTA, GGTA and TCTA motifs frequently found associated with mitochondrial promoter sequences (Kühn *et al.* 2005) are shown in bold. Numbers of clones sequenced for each site are specified by the frequency of their respective transcript 5' end as determined from TAP-treated (+TAP) or untreated (-TAP) RNA. 5' end frequencies are shown in 5'-to-3' order, if multiple TSS were found. Mutant 5' ends of *AtrpoTmp* plants were analysed from rpoTmp-2 5'-RACE products, but if rpoTmp-1 products formed a clearer band, the latter were cloned and sequenced (numbers marked #). P*cox1*-355 has been characterised previously. Therefore wild-type RACE products were not sequenced. For reported poorly defined *matR* 5' ends (Forner *et al.* 2007) sites are marked by asterisks. The first codons of the annotated *matR* and *rps4* coding sequences are italicised. –, Signal absent in 5'-RACE; n.a., not analysed. a, determined previously (Kühn *et al.* 2005). b, determined previously (Forner *et al.* 2007). c, determined previously (Forner *et al.* 2008) with minor differences. e, putative promoters (insufficient experimental data to support function as TSS).

In Figure 28, a graphical summary of TSSs determined in this study using total cellular RNA prepared from 7-d-old seedlings is provided. As exemplified by 5'-RACE products from cox1 transcript 5' termini (Fig. 29), the majority of 5' ends detected in the wild type were also detected in rpoTmp mutant plants (Fig. 29, Tab. 17). For some promoters, such as P*nad2e1*-114 and P*nad2e1*-297, sequence alignment of cloned primary transcripts revealed also 5' ends that did not map to exactly the same nucleotides in rpoTmp (lowercase positions in Tab. 17) and the wild type (underlined positions). An interpretation that different starter nucleotides reflect genotype-specific differences between initiating nucleotides in *AtrpoTmp* and the wild type has to be refuted since for some TSSs such differences were also observable between two independent 5'-RACE analyses performed on the same gene in the wild type (data not shown). In turn it seems that technical inconsistency of the 5'-RACE technique is more likely to account for this observation.

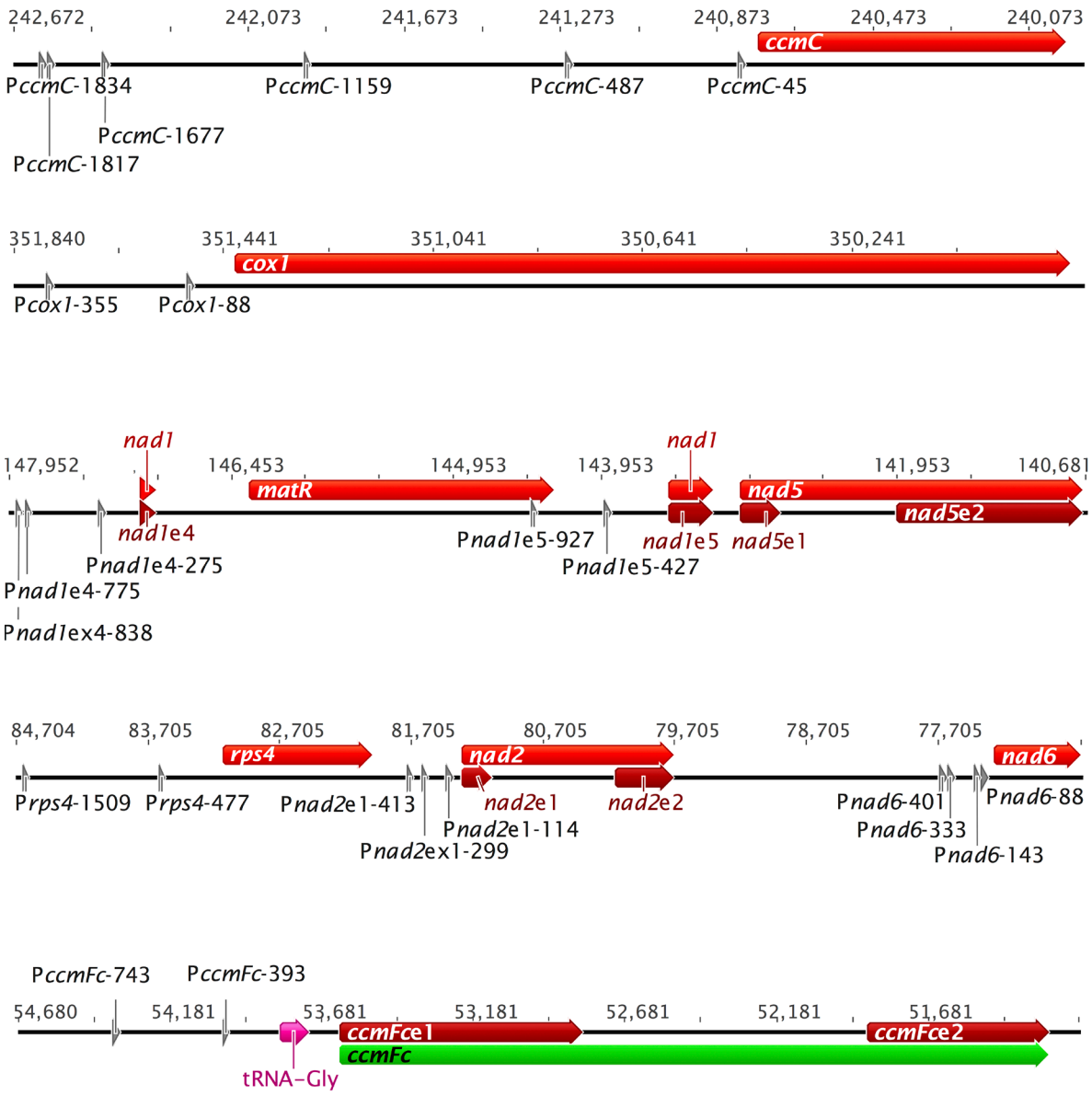


Figure 28: Maps of *Arabidopsis* mtDNA regions for which TSSs were analysed. TSSs as mapped in 5'-RACE analyses are indicated by grey triangles; genes with elevated or unchanged accumulation are shown as green arrows, whereas genes with lowered transcription are shown in red. For *cis*- and *trans*-spliced transcripts, exons are labeled with the gene name, followed by the exon number (e.g. nad1e4). Numbers indicate coordinates of the analysed regions on the mtDNA.

While selected transcript 5' termini, such as *nad6*-179, were barely detectable in mutants (Fig. 29), a primary 5' end mapping to P*nad6*-143 downstream of the fore mentioned processing site was more easily detected in rpoTmp plants (Fig. 29). Products mapping to P*nad6*-333 were neither detected in the wild type nor in rpoTmp-1 but in rpoTmp-2. These discrepancies are likely due to the very low abundance of the P*nad6*-333 5' end and can therefore not be regarded as genotypical differences but rather have to be interpreted as experimental variations.

Surprisingly, the data provided by the extensive analysis of TSS in *AtrpoTmp* plants did not indicate lacking transcriptional initiation from a specific subset of promoters preceding the genes for which the reduction of transcript accumulation had been observed. To the contrary

all promoters identified in the wild type were also found to be active in mutant plants. As a reminder it should be noted here that as in *AtrpoTmp* plants no changes in the utilisation of organellar promoters had been observed between *Physcomitrella* wild-type and *PprpoTmp2* plants (see 3.1.4).

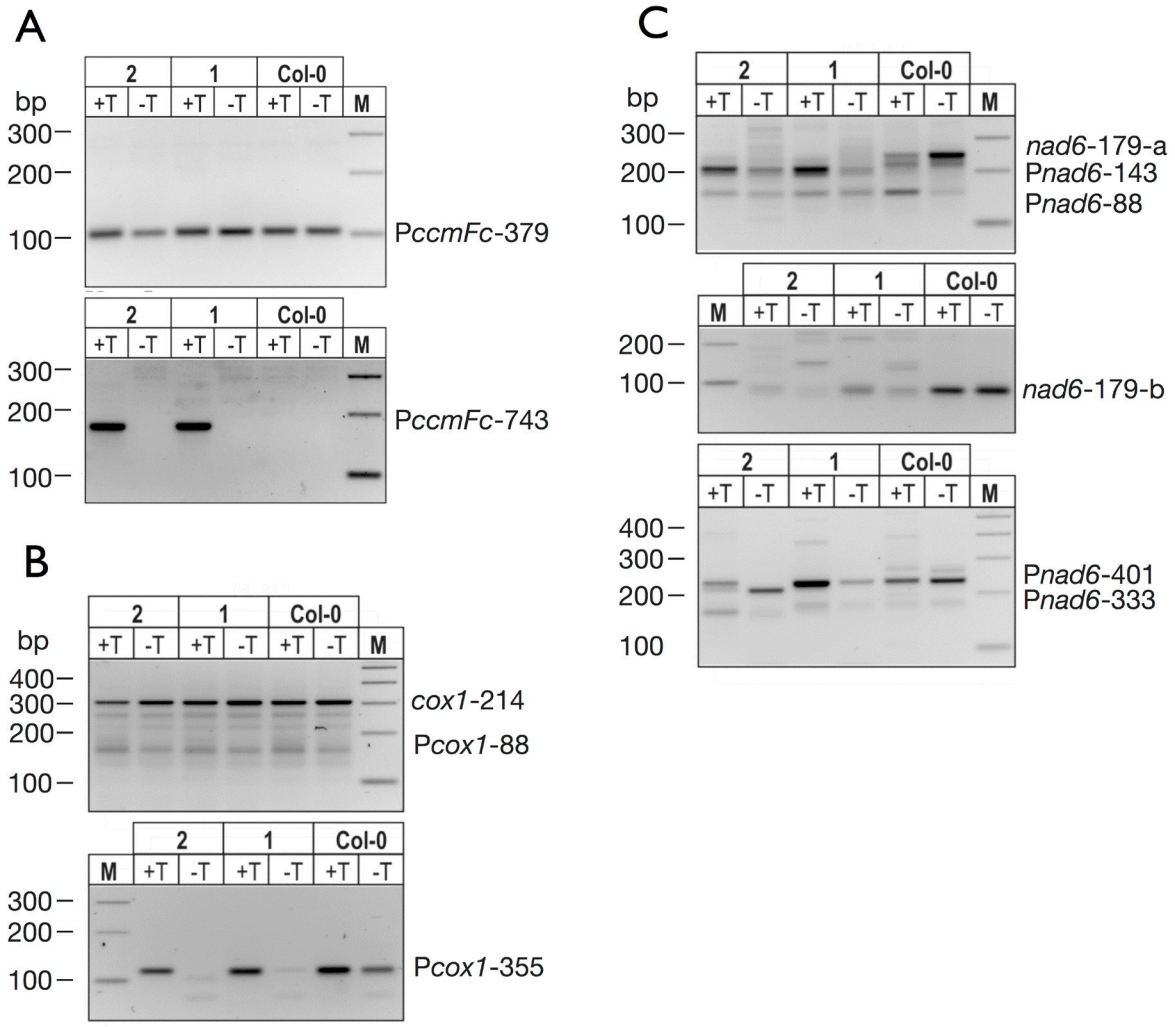


Figure 29: 5'-RACE analysis of *ccmFc*, *nad6* and *cox1* transcripts. 5'-RACE performed on RNA extracted from 7-d-old rpoTmp-1 (lane 1), rpoTmp-2 (2), and wild type (Col-0). PCR products from TAP-treated (+T) and untreated RNA (-T) reactions were separated on agarose gels alongside a molecular weight marker (M). Signals indicating primary transcript 5' ends are labeled with the name of the promoter listed in Table 17. Identified processed 5' ends are labeled as well. Sequencing of *ccmFc* transcripts mapping to position -379 was insufficient to support these transcripts being primary. Nevertheless sequence inspection revealed a consensus tetranucleotid directly upstream of the 5' end from both +TAP and –TAP reactions. Therefore P*ccmFc*-379 was considered a putative promoter. P*ccmFc*-743 seems to be inactive in the wild type but is consistently detected in mutants. While the product from the processed 5' end *nad6*-179 using reverse primer P4nad6 was barely detectable in the mutant (signal marked *nad6*-179-a), amplification with a the reverse primer P5nad6 further upstream to *nad6*-179 not detecting other downstream 5' ends (*nad6*-179-b) gave rise to weak but detectable signals. The *cox1* promoter P*cox1*-355 has been characterised previously; no detailed mapping was therefore done on wild-type transcripts mapping to this site. All other data from 5'-RACE analysis as listed in Table 17 are shown in Suppl. Figure 5.

Interestingly, primary 5' ends mapping to P*ccmFc*-743 were consistently detected in mutants but not in the wild type, indicating that P*ccmFc*-743 is inactive in the wild type. As in mitochondria of *Physcomitrella rpoTmp2* plants, where additional promoters were observed

(see Fig. 15) *ccmFc*-743 is found consistently only in *AtrpoTmp* plants (Fig. 29A). While intuitively a promoter used in the absence of AtRpoTmp should still be used when the enzyme is active, especially high levels of the *ccmFc* transcript (see Fig. 26) might partially be driven by the additional use of P*ccmFc*-743. Without assuming lack of transcriptional initiation at *ccmFc*-743 in the wild type, rapid processing of the corresponding 5' transcript end in wild-type plants could also account for this observation.

Since lacking initiation from certain mitochondrial promoters was not the reason for the observed reduction of transcript accumulation in mitochondria of plants lacking AtpoTmp, it was speculated that AtRpoTm might be able to substitute for AtRpoTmp at promoters of a certain architecture but with lesser efficiency. Therefore, Dr. Kühn attempted to determine the abundances of these 5' termini in mutants and the wild type using quantitative ribonuclease protection assays (RPAs). The quantitative analysis of transcript 5' ends showed that all 5' termini of a transcript with reduced transcript accumulation were diminished in the mutant (see Suppl. Fig. 8). Since lowered abundances of subgenomic DNA molecules in rpoTmp were not observed (see Suppl. Fig. 9), promoter mapping together with run-on transcription and RPA data indicate that reduced transcriptional initiation on multiple and diverse promoters, preceding the genes with lowered transcript accumulation, are likely the direct cause for the reduction of respiratory complexes I and IV in the mutant.

3.2.2.3 Promoter structure analysis of genes with reduced transcription

Figure 30 shows a list of all promoters and putative promoters identified in *Arabidopsis* mitochondria to date. Sequences spanning nucleotides, from position -14 to +10 around the TSS, are displayed. The region from -14 to +4 has been defined by Dombrowski et al. (1999) to be required for promoter function in dicot plant mitochondria (Dombrowski *et al.* 1999). Furthermore for T7-like phages it has been shown that the initially transcribed 10 nucleotides are important for a mechanism described as abortive cycling, which reduces the switch to the more stable elongation phase and thereby effects the efficiency of a promoter. A comparison of promoters and putative promoters of genes from which transcription initiation was shown to be reduced in *AtrpoTmp*, and all other known mitochondrial promoters in *Arabidopsis* did not detect any common motif, that would identify these promoters as a distinct subset of mitochondrial promoters. The only sequence element found to be statistically overrepresented amongst promoters conferring reduced transcriptional activity in the mutant is the GAAA upstream tetranucleotide, also found to precede mitochondrial promoters mapped in *Physcomitrella patens*.

Since not all of the promoters of genes transcribed less actively in the *Arabidopsis* rpoTmp mutant contain this element and the positioning is rather variable (Fig. 30), this element alone is unlikely to confer RpoTmp-specific promoter recognition or repression of the ability of AtRpoTm to effectively initiate transcription when substituting for RpoTmp in the mutant. Thus the only criterion by which these promoters are defined as a distinct subset of mitochondrial promoters was the fact that transcription initiation was reduced in *AtrpoTmp* plants. As a result control of mitochondrial transcription in the eudicot plant *Arabidopsis thaliana* might be best described as partially conferred through elements within the transcribed regions rather than by sequence elements upstream of the TSS.

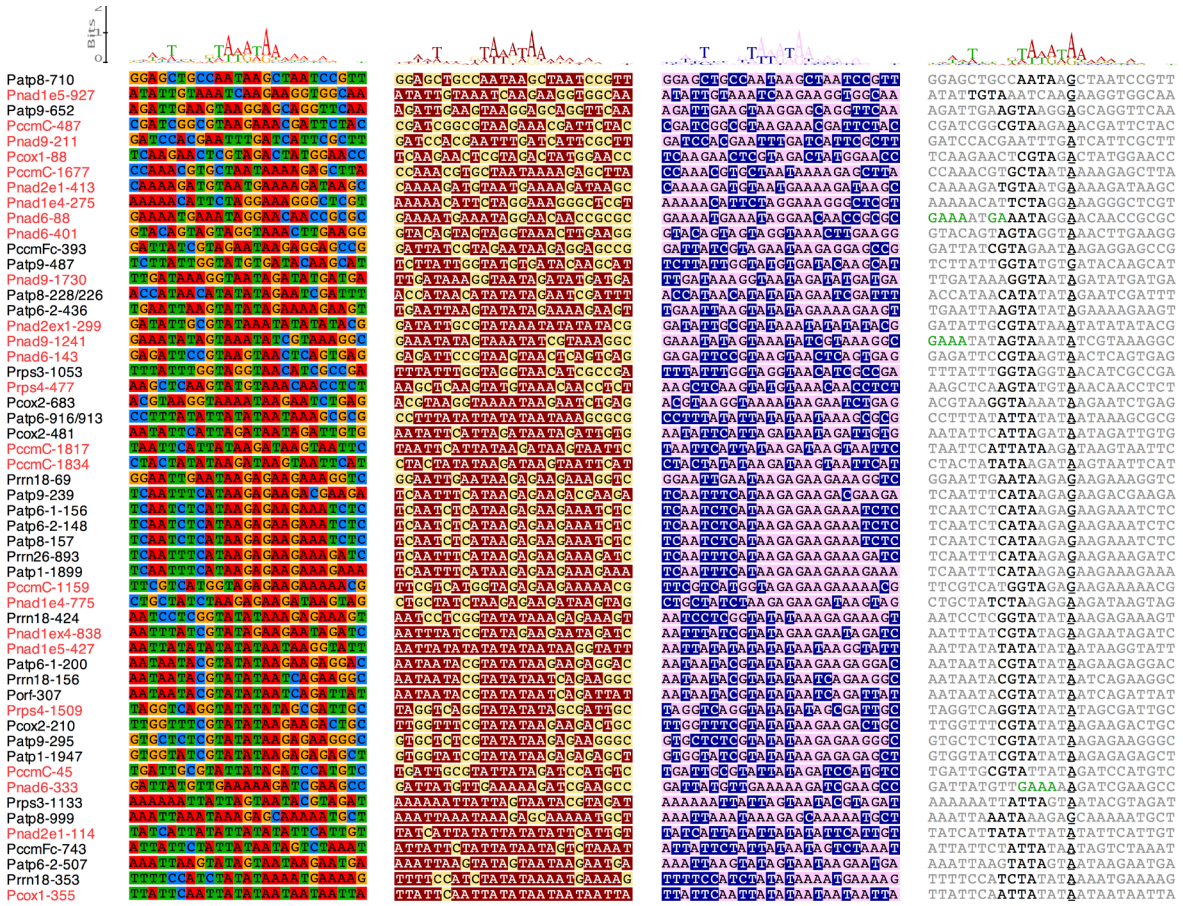


Figure 30: Promoter Sequences in Arabidopsis Mitochondria. The ClustalW (Thompson *et al.* 1994) alignment includes promoters identified previously (Kühn *et al.* 2005) and in this study. The names of promoters of genes showing reduced transcription in rpoTmp mutants are printed red. Far right: Tetranucleotide motives as CATA, CGTA, TCTA, ATTA, AGTA, or GGTA immediately upstream of the TSS (underlined) are indicated in bold black. The tetranucleotide GAAA upstream from the TSS, only found in the subset of promoters conferring reduced transcription, is printed green. Identity (far- left), purin - pyrimidin (mid-right) and AT-GC (mid-left) coloring was used to observe patterns common to promoters with reduced transcriptional activity. Neither of the matrices is indicative of a common motive dividing those promoters from the promoters of none-reduced transcripts.

3.3 On the evolution of phage-type RNA polymerases

3.3.1 On the origin of eukaryotic phage-type RNA polymerases

Transcription is catalysed by two distinct families of DNA-dependent RNA polymerases. RNAPs with a remarkably conserved multi-subunit structure have been found to operate in all cellular organisms including several polymerases acting on DNA viruses, eukaryotic enzymes Pol I, Pol II, Pol III and Pol IV and their archaeal, bacterial and chloroplast counterparts. The other family, the monomeric or single subunit RNA polymerases (ssRNAPs) belong to a super-family of ‘right-handed’ DNA and RNA polymerases (Superfamily A) and are involved in the transcription of mitochondrial, chloroplast and phage genomes (Cermakian *et al.* 1997, Greenleaf *et al.* 1986, Hedtke *et al.* 1997, Masters *et al.* 1987, McAllister and Raskin 1993, Weihe *et al.* 1997).

To shed light on the evolution of plant phage-type RNA polymerases and on monomeric DNA dependent RNAPs in general a comprehensive phylogenetic analysis was carried out, using the data produced by our laboratory (Yin *et al.* 2009, Yin *et al.* 2010) and the steadily growing information from the public databases. The database search employing the BlastP algorithm via NCBI blast search webpage to date gives over 230 hits for putative homologs of ssRNAPs. Due to the reported trichotomy of the ssRNAP cluster (Cermakian *et al.* 1997), searches employed three sequences from each subgroup, namely ssRNAP sequences of *Selaginella moellendorfii* (nuclear-eukaryotic), *phiKMV* (Podoviridae; NP_877465) and *Daucus carota* (mitochondrial plasmids; AAS15052).

The most conserved blocks in the derived alignment (Suppl.Fig. 6) using COBALT, a **Co**nstraint-**b**ased Multiple **Al**ignment **T**ool (Papadopoulos and Agarwala 2007) represent parts of the regions delineated by Masters et al. (Masters *et al.* 1987). The N-terminal region is not well conserved in ssRNAPs (Cermakian *et al.* 1997). Even in otherwise highly similar plant nuclear encoded ssRNAPs the N-termini show significant deviation, indicating that the N-termini evolve rapidly. In this respect it is important to note that the N-terminal part of the T7 polymerase has been shown to harbour motives important in promoter binding, whereas catalytically relevant structures of the polymerase are formed through amino acids of the C-terminal half of the enzyme.

Though the derived alignment is in favour of a common origin of ssRNAPs, a pattern of additional modular evolution is suggested by two observations. First, two PPR domains are found in mammalian nuclear encoded mitochondrial RNAPs (Ringel *et al.* 2011) and second, insertions between the blocks designated K and L, as found in fungal polypeptides (Cermakian *et al.* 1997), are lineage specific 'inventions', that possibly accomplish specific functions of ssRNAPs in these taxa. Neither of these insertions are found in any of the newly aligned sequences, but in vertebrates and fungi, respectively (data not shown).

In general 170 eukaryotic proteins with high similarity to nuclear encoded RpoTs were included in the alignment. Additional blast hits were excluded either because the annotated sequences were truncated or because they showed clear signs of partially false ORF predictions. From blast searches with the phiKMV protein 50 phage blast hits were taken, excluding truncated proteins. Additionally 15 bacterial blast hits were identified. In all cases of 'bacterial' blast hits inspection of the encoding genomic locus revealed other phage-like ORFs suggesting, that all of the ssRNAP ORFs are part of prophage insertions or remnants of such, as has been shown for *Pseudomonas putida* KT2440 (Shutt and Gray 2006a) and *Agrobacterium tumefaciens* (Filee and Forterre 2005). Moreover, though ssRNAP family ORFs are found in prokaryotic genomic sequences, all those genomes were also found to encode the typical multi-subunit RNAPs. In supplementary Table 1 a list of all proteins used for the phylogenetic analysis is given.

Phages containing a single enzyme RNAP of the T3/T7 type all belong to the subfamily of *Autographivirinae* except for four *Xanthomonas* infecting phages (*Siphoviridae*), three unclassified Vibrio phages, three uncultured phages (MedDCM-OCT isolates) and two unassigned caudoviral Cyanophages. The host specificity of all phages appears to be restricted to proteobacteria except for the four cyanobacteria infecting phages (Cyanophage 9515-10a, Cyanophage Syn26, Synechococcus phage P60 and Synechococcus phage Syn5). In turn it is of no surprise, that all blast hits from bacterial genomes show the same taxonomic distribution with the only exception of an ORF identified in *Pyramidobacter piscolens*, a member of the bacterial phylum Synergistetes.

Phylogenetic trees were constructed from an alignment using 235 taxa and from highly conserved blocks (Suppl. Fig.6) extracted from the fore-mentioned Cobalt-alignment. For a phylogenetic reconstruction based on Bayesian statistics, the MRBAYES program version 3.1 (Ronquist and Huelsenbeck 2003) was used. In addition, maximum likelihood (Guindon and Gascuel 2003) analysis implemented in the GENEIOUS package (Drummond *et al.* 2008) was conducted. Both analyses yielded the same overall topology shown in Figure 31, with only two well-defined clusters. While in an earlier analysis (Cermakian *et al.* 1997), three clusters and in turn three classes of ssRNAPs were defined. The unrooted tree in Figure 31 only shows two clearly separated groups, due to the nesting of the plasmid-encoded proteins inside the phage cluster. The plasmid-encoded sequences are nevertheless likely monophyletic, as are the nucleus-encoded ones. The identification of four multi-subunit eubacterial-like *rpo* genes in the mtDNA of the early diverging eukaryote *Reclinomonas americana*, lends strong credence to the view that mitochondrial transcription initially employed the α-proteobacterial multi-subunit RNAP. This observation together with the distribution of ssRNAP in proteobacteria and the phylogenetic analysis presented herein strengthens the view of an origin for nuclear encoded ssRNAPs from prophage insertions in the genome of the ancestral α-proteobacterial endosymbiont. However, because the phylogenetic tree is unrooted, the order of the divergence cannot be inferred from the available data. Nevertheless the positioning of the α-proteobacterial *Rhodopseudomonas* phage-like sequence in the tree in sister position to all other phage sequences and to the eukaryotic nuclear encoded proteins is intriguing.

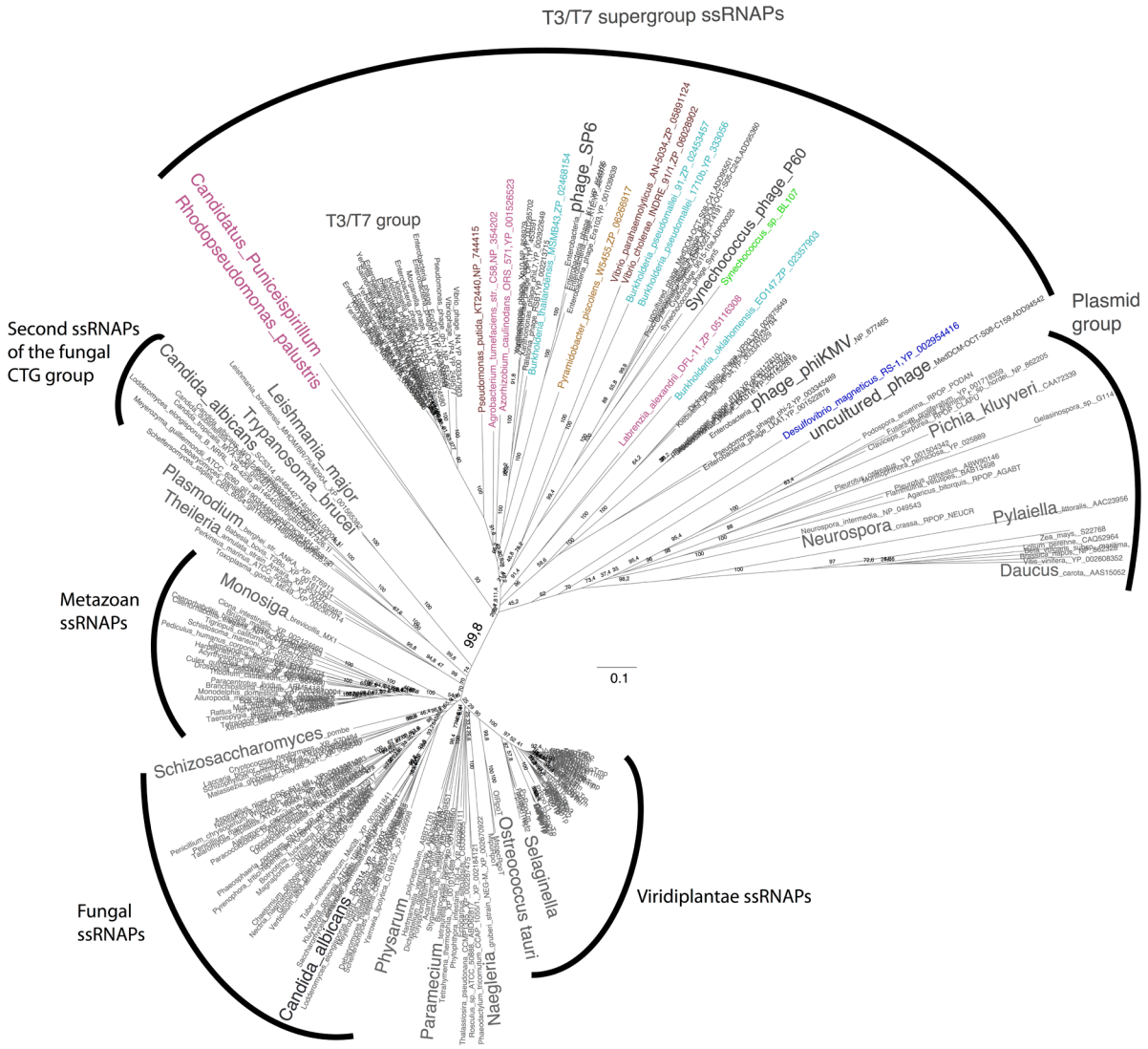


Figure 31: Phylogeny of monomeric DNA dependent RNA polymerases. The phylogeny was reconstructed using PhyML from conserved amino acid sequence sections indicated in the amino acid sequence alignment (see Suppl. Figure 6). Clades are marked with higher order taxon names and representative model species are shown enlarged. All proteins derived from monomeric RNAP encoding ORFs found in bacterial species are coloured as follows: magenta for α-proteobacterial, pale blue for β-proteobacterial, brown for γ-proteobacterial dark blue for δ-proteobacterial and green for cyanobacterial proteins. The protein as found to be encoded in the Synergistetes species is shown in pale brown. The phyML evolutionary tree represents members of all the three previously described supergroups (eukaryotic nuclear, mtORF and phage proteins), the relationships between which remain unresolved due to the unrooted nature of the tree. Nevertheless the mitochondrial ORF encoded polymerases are now nested inside the phage cluster. Numbers at branching points are branch support values. Branch lengths correspond to the number of inferred amino acid changes per position, as indicated by the scale bar. The phylogeny was reconstructed by the PhyML algorithm V3.0 (Guindon and Gascuel 2003)implemented in the Geneious v5 package (Drummond *et al.* 2010)with 1000 bootstrap replicates. The aligned 235 sequences comprised 299 amino acids. For the reconstruction the wag amino acid replacement model was employed. A discrete gamma distribution with four categories was assumed to approximate the continuous function (shape: 0,934/ inv: 0,02).

Thus far multiple ssRNAPs were only observed in land plants. Therefore one fascinating additional observation was the identification of eight fungal species each encoding two putative phage-type RNA Polymerases in their nuclear genomes.

By inspection of the amino-acid alignment it was obvious that the two annotated proteins from each of the eight species where rather distantly related, with one homolog clearly representing the 'usual' fungal ssRNAP lineage, whereas the other putative ssRNAP protein is distantly related to the other fungal sequences but still showed higher similarity to the nuclear, then to mitochondrial-plasmid-encoded or viral polypeptides (data not shown).

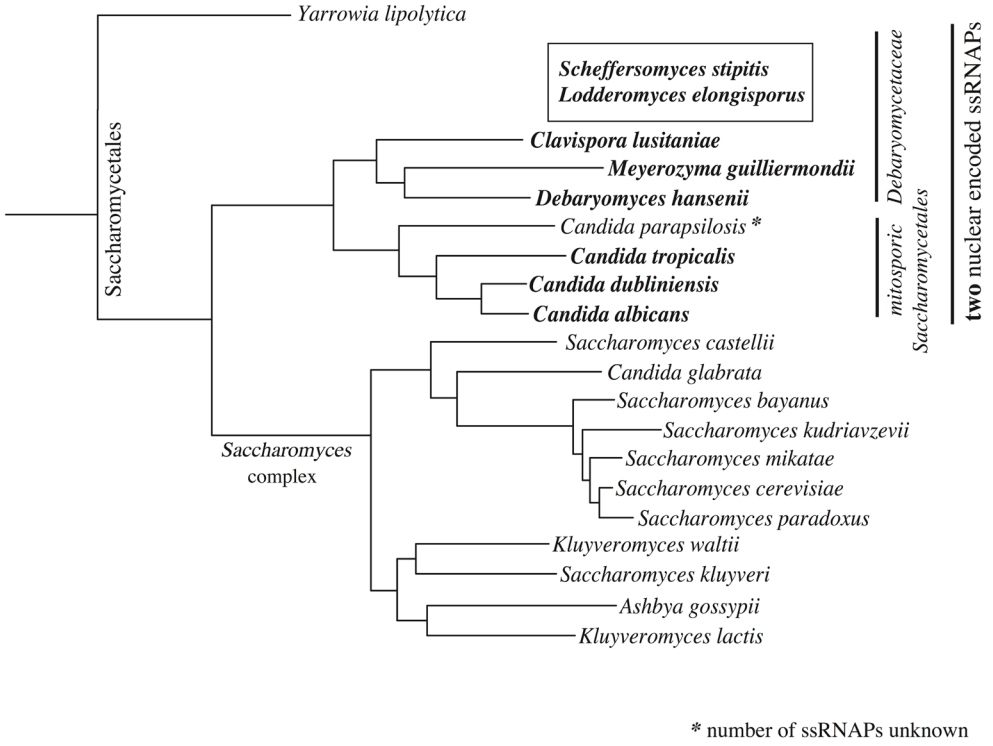


Figure 32: Taxonomy of Saccharomycetales and distribution of a second putative ssRNAP encoded in fungal nuclear genomes. Two putative ssRNAPs are found in eight Saccharomycetales species. *Debaryomyces hansenii, Meyerozyma guilliermondii, Clavispora lusitaniae, Scheffersomyces stipitis* and *Lodderomyces elongisporus* belong to the family of Debaryomycetaceae, whereas *Candida albicans, Candida dubliniensis* and *Candida tropicalis* are species belonging to mitosporic Saccharomycetales. Figure was adapted from Scannell et al. (Scannell *et al.* 2007).

The positioning of the second ssRNAP in the phylogenetic tree confirmed the observation, that all phage-type RNAP proteins of the eight fungal species belong to the cluster of nuclear encoded mitochondrial ssRNAPs (see Fig. 31). While the 'usual' fungal polypeptides are clearly affiliated with the nuclear enconded fungal mitochondrial RNAPs as of *Saccharomyces cerevisiae*, the second homolog surprisingly appears as a deep branching eukaryotic nuclear encoded mitochondrial RNAP (see Fig. 31). *Debaryomyces hansenii, Meyerozyma guilliermondii, Clavispora lusitaniae, Scheffersomyces stipitis* and *Lodderomyces elongisporus* belong to the family of Debaryomycetaceae, whereas *Candida albicans, Candida dubliniensis* and *Candida tropicalis* are species belonging to mitosporic Saccharomycetales. The distribution of the additional ssRNAP indicated in Figure 32 suggests acquisition in a shared common ancestor of both families, but after the split from the ancestor of the Saccharomyces complex. Taken together these data suggest, that multiple phage-type RNA polymerases might not be a unique feature of land plant species and that the second putative ssRNAP in fungal species of the Saccharomycetales sub-cluster was acquired by a single horizontal gene transfer (HGT) event.

3.3.2 Evolution of RpoT proteins in Viridiplantae

Using the Bayesian algorithm as well as maximum-likelihood (PhyML) analysis, phylogenetic trees were reconstructed to revaluate the molecular phylogeny of the RpoT polymerases in the plant kingdom and more specifically to reassess the phylogenetic positioning of the polymerases characterised in the present study. Tree reconstruction was based on a multiple alignment of 55 RpoT sequences (see Fig. 33; Suppl.Tab. 1). Bayesian as well as PhyML analysis resulted in essentially the same tree topology (not shown). Trees were rooted with the RpoT proteins from green algae. Figure 33 shows the rooted consensus tree of the PhyML analysis in which angiosperm RpoT polymerases constitute two clearly discernible groups: one consisting of plastid localised polymerases, and the other of mitochondrial-localised and dual-targeted enzymes. Whereas *Pinus*, *Selaginella* and *Physcomitrella* polymerases do not belong to the branches of well separated plastid and mitochondrial (and dual targeted) polymerases, the RpoT polymerases from the basal angiosperm *Nuphar advena* cluster with the branches of plastid and mitochondrial/dual targeted sequences: NaRpoTm1 and NaRpoTm2 within the mitochondrial, and NaRpoTp within the plastid branch. The positioning of the RpoT protein from *Pinus taeda* with putative mitochondrial targeting properties (Beick et al., unpublished data) on the tree is in agreement with the land plant phylogeny.

The *Selaginella* protein SmRpoT is neither affiliated with the branches of the well-separated plastid and mitochondrial (and dual-targeted) polymerases nor is the positioning on the tree congruent with the well established land plant phylogeny. Instead, SmRpoT rather appears as a paralog of all other land plant RpoT proteins including mosses. Since bootstrap support is sufficiently high, one has to assume an ancestral state of at least two RpoT proteins in the common ancestor of bryophytes and later branching land plants. Accordingly, the orthologue of the *Selaginella* RpoT protein paralogous to all other land plant RpoTs was lost twice independently in the evolution of land plants: firstly in the *Physcomitrella* lineage and secondly after the split of the *Selaginella* lineage from all higher branching tracheophytes. The orthologue of all other land plant RpoT proteins was lost in the *Selaginella* lineage. All *Physcomitrella* RpoTs on the other hand appear to be the result of two *Physcomitrella* lineage specific duplication events, with the PpRpoTm/RpoTmp1 duplication as the more recent one.

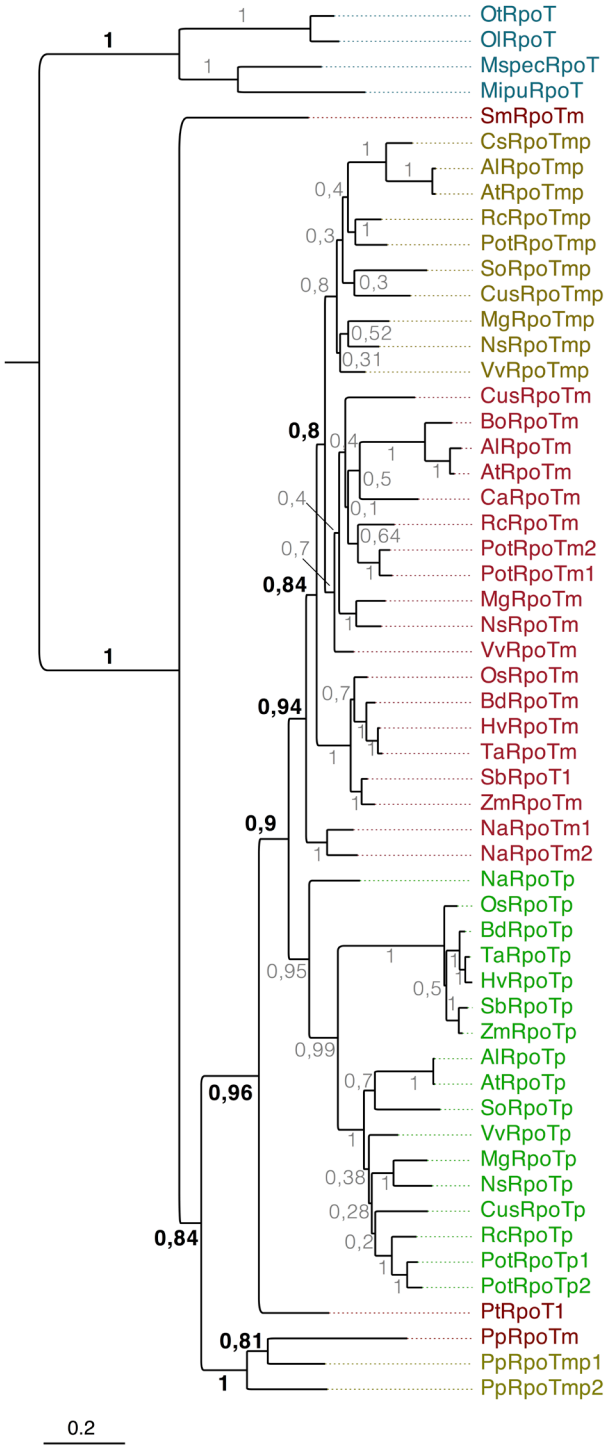


Figure 33: Phylogenetic analysis of plant RpoT proteins. The phylogeny was reconstructed by the PhyML algorithm V3.0 (Guindon and Gascuel 2003) implemented in the Geneious v5 package (Drummond *et al.* 2010) with 100 bootstrap replicates. The aligned sequence data comprised 894 amino acids lacking only the less conserved N-terminal portion of the proteins. For the reconstruction the wag amino acid replacement model was employed. A discrete gamma distribution with four categories was assumed to approximate the continuous function (shape: 1,004/ inv: 0,138). Prefixes of designations of plant RpoT proteins refer to Abbr. in Suppl. Table 1. Essentially the same tree topology was obtained employing Bayesian (Ronquist and Huelsenbeck 2003) and neighbor joining analysis (Saitou and Nei 1987). Important bootstrap support values are indicated in bold. The tree was rooted with the four algal (green algae) proteins indicated in blue (m designates mitochondrial enzyme, p plastid enzyme, and mp enzyme with dual targeting).

4 Discussion

4.1 Identification of transcription initiation sites

As outlined before *in vitro* capping, *in vitro* transcription assays, the *RACE-TAP* based method and *in silico* sequence analysis has led to the identification and prediction of mitochondrial and plastid promoter motifs and transcription initiation sites in a variety of mono- and dicotyledonous plants, as well as in green algae (Liere *et al.* 2011).

Mitochondrial genomes except those of *Jacobides* lost the genes that encode the multi-subunit bacterial-type RNA polymerase. Therefore, transcription in mitochondria is strictly dependent on nuclear encoded RNAPs. In concordance with the cyanobacterial nature of plastids and the observed retention of the genes encoding a multi-subunit RNAP of the bacterial-type in the plastomes of almost all land plants, plastid transcription is initiated from bacterial-type promoters (PEP promoters) in the plant lineage. Additionally NEP I and NEP II sequence motifs have been shown to drive plastid transcription in flowering plants (Liere *et al.* 2011).

Albeit the mitochondrial and plastid genomes of the deep branching land plants were among the first to be sequenced (Oda *et al.* 1992, Ohyama *et al.* 1986), no experimentally verified transcription initiation sites and/or organellar promoters have been described to date in deep branching land plants like *Physcomitrella*. Therefore, this study provides the first data on plastid and mitochondrial transcription initiation in a bryophyte species. Even though lack of organellar transcriptional initiation in *PprpoTmp2* plants as compared to the wild type was not observed, eleven mitochondrial and nine plastid transcription initiation sites were mapped.

4.1.1 Plastid promoters in *Physcomitrella patens*

Genes exclusively transcribed from NEP promoters found to date are rare, including *accD*, *ycf2* in dicots and the *rpoB* operon in flowering plants.

The mapping of TSS for *Physcomitrella psaA* (PpP*psaA*-161), *psbA* (PpP*psbA*-54), *atpB* (PpP*atpB*-270, PpP*atpB*-214, PpP*atpB*-144), *accD* (P*accD*-42, P*accD*-36), *trnG* (PpP*trnG*-32) *ycf2* and *rpoB* (P*rpoB*-28) revealed all identified TSS to be preceded by motifs with similarity to described PEP promoters. The TSS identified herein were found to be preceded by -10 regions typically observed in promoters recognised by the bacterial-type plastid polymerase. Four out of nine plastid TSS lack the typical -35 region, including P*accD*-42, P*accD*-36, PpP*atpB*-214 and P*rpoB*-28. This lack of the -35 element is reported for a variety of PEP promoters in flowering plants (Liere *et al.* 2011) and *Chlamydomonas* (Klein *et al.* 1992). Moreover, the lack of a typical conserved -35 box is also observed in bacterial promoters that are subject to positive control, meaning that additional cofactors are needed for full 'activity'. Plant candidate proteins to assist transcriptional initiation on plastidial promoters lacking the -35 region are the well characterised sigma-factors, which in analogy with eubacterial counterpart sigma-70 family proteins, define binding specificity to target promoter sequences (Allison 2000) and have also been discovered in *Physcomitrella patens* (Hara *et al.* 2001, Ichikawa *et al.* 2004).

Taken together, the plastid promoter architecture in *Physcomitrella* seems to recapitulate structures observed for plastid promoters identified in other land plants. The -10 region is located from -13 to -10 with respect to the TSS and there is remote variation in the length of the spacer region between the -35 and -10 box (+/-3).

In *Arabidopsis thaliana* and tobacco promoters of *ycf2* are found in great distance to the *ycf2-orf* (AtP*ycf2*-1629, AtP*ycf2*-1506, NtaP*ycf2*-1577). When the TSS for Pp*ycf2* was approached the only bona fide primary 5' end was mapping to the upstream region of *psbA* located upstream of the *ycf2* and *trnaH* coding sequences. The mapped TSS in *Physcomitrella* is located 1627 nucleotides upstream with respect to the *ycf2* translational start site.

Therefore, it is likely that in *Physcomitrella* protonemata *ycf2* is dominantly cotranscribed with *psbA* and *trnaH* from the *psbA* promoter P*psbA*-54. Tough the extended 5'UTR seems to be a common feature between tobacco, *Arabidopsis* and *Physcomitrella*, *ycf2*-cotranscription of upstream ORFs is not observed in *Arabidopsis* and tobacco since coding capacity is located only on the opposite strand in those species. One peculiarity of some bryopside plastid genomes is the lack of *rpoA* coding capacity (Goffinet *et al.* 2005). Different to other photosynthetically active land plants the *rpoA* gene was transferred to the nuclear genome in *Physcomitrella*. In the nuclear genome sequence of *P. patens* two genes encode two RpoA homologs (Kabeya *et al.* 2007). Since the RpoA subunit of PEP is thought to facilitate interaction of the core subunits of PEP with sigma factors, it is tempting to speculate that different combinations of the two RpoA proteins and the three sigma factors are involved in the regulation of transcription at promoters with varying structures (e.g., lacking the -35 box).

In flowering plants the *rpoB* operon encodes three out of four PEP core subunits. The transcription of this operon is reported to be transcribed by NEP only (Hajdukiewicz *et al.* 1997, Hübschmann and Börner 1998, Silhavy and Maliga 1998, Swiatecka-Hagenbruch *et al.* 2007). The mapping of the *Physcomitrella rpoB* promoter revealed a major 5' end in very close proximity to the coding region, but instead of a conserved NEP promoter motif the initiating nucleotide was found to be downstream from a bacterial like -10 box typically found in PEP promoters. Moreover the initiating nucleotide is a thymidine as described for a variety of PEP promoters, whereas for the nuclear encoded phage-type RNA polymerases transcription initiation is highly biased towards A and G nucleotides.

Taken together, all plastid TSS mapped in protonemata 'tissue' seem to be preceded by PEP like promoter motifs, even *rpoB*, *accD* and *ycf2*. Though PpP*atpB*-144 was mapped as a promoter after stringent *in silico* search for NEP promoters, both NEP and PEP motifs seem to precede the transcription initiation site of PpP*atpB*-144.

Because NEP transcripts are usually, not detectable in green chloroplasts (Hess and Börner 1999, Liere and Maliga 2001) inhibition of plastid translation has been widely used to reveal NEP transcription activity and related promoters (Allison *et al.* 1996, Hajdukiewicz *et al.* 1997, Hübschmann and Börner 1998, Miyagi *et al.* 1998, Swiatecka-Hagenbruch *et al.* 2007). Unfortunately, spectinomycin treatment together with glucose supplementation of wild-type and also *PprpoTmp2* protonemata was lethal. It is difficult to infer any conclusion from this finding since addition of 3% glucose (0.17M) to the medium already had an inhibitory effect on the protonema culture.

In all promoter mapping experiments RNA isolated from protonemata was used and it can therefore not be excluded that in other developmental phases, such as the sporophyte plastid transcription is dependent on NEP promoters. A lack of NEP promoter activity in protonemata is in concordance with the following observations. Firstly, plastid transcription in protonemata of *Physcomitrella* is suggested to be independent of NEP since GFP targeting studies and immuno-blot data, suggest exclusive mitochondrial localisation of PpRpoTmp1 and PpRpoTmp2 in moss protonemata (Kabeya and Sato 2005). The second line of evidence

comes from the specific inhibition of the PEP by tagetitoxin, which reportedly led to complete loss of transcriptional activity in protonematal plastids (Kabeya *et al.* 2002). To the contrary tagetitoxin inhibition experiments in *PpRpoA* disruptant plants suggest, that tagetitoxin-resistant NEP activity might exists in the moss chloroplast (pers. communication Dr. Kobayashi, Nagoya University). Taken together, data provided herein support the view of a predominant role of PEP for plastid transcription in protonemata of *Physcomitrella*.

4.1.2 Sequences upstream of mitochondrial TSS exhibit typical features of plant mitochondrial promoters

When mitochondrial genome sequences of angiosperm species became available, the few experimentally defined promoter sequence motifs were used to screen for other promoters and therefore led to a biased picture concerning the diversity of sequences identified to drive mitochondrial transcription. This is in part also true for the groundbreaking work on mitochondrial transcription initiation in the model plant *Arabidopsis thaliana* (Kühn *et al.* 2005). However, a variety of mitochondrial promoters from several flowering plants have been described (see Tab. 3). The majority of flowering plant promoters identified to date reveal the existence of a conserved 5'-CRTA-3' core promoter element upstream and in close proximity usually 2 to 4 nucleotides from the initiation site (see Tab. 3). Six out of nine mitochondrial promoters identified in *Physcomitrella* in this work follow this rule (Fig. 34). More specifically five out of those six contain the 5'-CATA-3' core motif, but no promoter contains the two classes of nonanucleotides described for some dicot promoters. Two out of nine were found to contain a 5'-DDTA-3' core motif, a core element also observed in mitochondrial promoters of flowering plants. A common feature of both monocots and dicots is that initiation of transcription is preferred at adenine nucleotides. This is clearly a feature shared by *P. patens*. All mapped TSS were found to start with an adenine nucleotide. As in dicots, purines in general appear to be highly frequent at positions +1 and +2 with respect to the TSS.

Sequences upstream from the mapped TSS are found to be A/T-rich. This A/T-richness has been described to be important for proper promoter function in both monocots and dicots (Dombrowski *et al.* 1999, Rapp *et al.* 1993, Rapp and Stern 1992) and it is a structural feature of mitochondrial promoters in fungi (Tracy and Stern 1995). In addition to formerly described patterns the alignment of all mitochondrial promoters mapped in wild-type *Physcomitrella* potentially indicates conservation of two more sequence elements. First, in a distance of -40 to the TSS an additional AT-rich sequence seems to be conserved. Even more surprising is the observed pattern around +25 to +30 with respect to the TSS, where conservation of two A nucleotides is observable. The alignment of downstream sequences from *Arabidopsis* mitochondrial TSS did not reveal any similar pattern. Mutagenesis of the newly identified motifs and subsequent *in vitro* transcription studies using such promoter constructs could clarify, if these motifs have any influence on promoter recognition or strength.

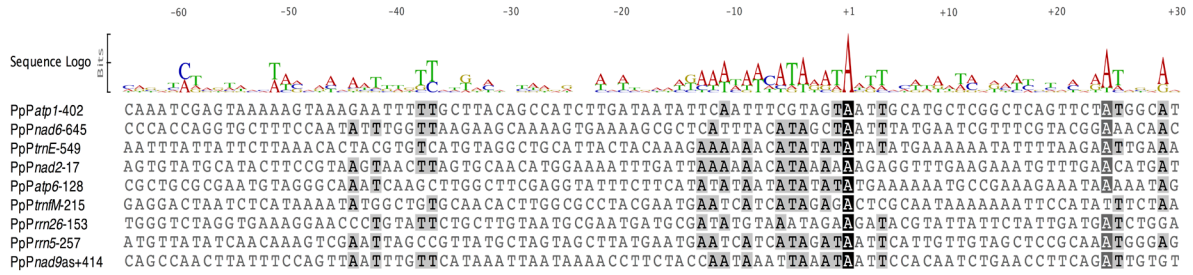


Figure 34: Mitochondrial promoters in *Physcomitrella patens*

In flowering plants a mitochondrial gene is often transcribed from multiple promoters. Lupold et al. suggested that this might be either a mere reflection of a promiscuous transcriptional apparatus or that multiple promoters are needed to counteract possible consequences of frequent intramolecular recombinations of the mitochondrial genome (Lupold *et al.* 1999). When TSS of the *rrn26* transcript were mapped in the wild-type moss, three promoters were found to drive transcription of *rrn26*; P*rrn5*-257, P*trnfM*-215 and P*rrn26*-153, suggesting that the 26S rRNA is transcribed from multiple promoters in *Physcomitrella*. It thus seems that this basic mechanism of mitochondrial transcription is common throughout land plants. Furthermore, mitochondrial core promoter motifs in *Physcomitrella* do not significantly deviate from the structure of mitochondrial promoters described for other land plants.

4.1.3 Antisense transcription from a mitochondrial promoter downstream of *nad9*

A transcription initiation site was detected in the present study that gives rise to an antisense transcript of *nad9* initiated from P*nad9*as+414 in *Physcomitrella*. This TSS is found directly upstream from the translational start site of *atp1*. Accordingly, transcription from P*nad9*as +414 produces a RNA, which is not only reverse complementary to the *nad9* transcript, but also to the 5'UTR of *atp1*. Moreover, inspection of the *nad9*as transcript revealed a 24 nucleotide motif, shown to be conserved at 5' processing sites of the mitochondrial transcripts *nad6*, *atp9* and *rrn26* from *Arabidopsis thaliana* (Forner *et al.* 2007). No functional data exist to date on the possible function of these motifs, nevertheless Forner et al. suggested, that these primary sequences might be important for the generation of processing 5' transcript ends. Considering the apparent conservation during 450 million years of land plant evolution, it would be of great interest to extensively map this putative processing site in *Physcomitrella patens* and to determine the functional significance of transcription and transcript accumulation of the *nad9*/*atp1* locus.

Antisense transcripts of known genes are reported from both plant organelles (Finnegan and Brown 1990, Holec *et al.* 2008b, Holec *et al.* 2006, Kühn *et al.* 2005, Sharwood *et al.* 2009). It was speculated that in case of mitochondria these RNAs might have no function but are instead the result of a relaxed transcriptional control and of non-stringent sequence requirements of plant mitochondrial promoters, indicated by the highly diverse promoter architecture and multiple promoters driving the transcription of the same gene indifferently (Holec *et al.* 2006, Kühn *et al.* 2005). However, the antisense promoter mapped in *Physomitrella* as well as seven out of nine of such antisense TSS, identified in *Arabidopsis* mitochondria, map to complex I coding regions (Kühn 2006).

Beside a major role of antisense RNAs in cyanobacterial gene regulation (Georg *et al.* 2009), in plastids it was shown, that a polyadenylated transcript from *Chlamydomonas* is stabilised by asRNA transcription from a convergent promoter (Nishimura *et al.* 2004), and that in tobacco an asRNA interferes with C-to-U editing (Hegeman *et al.* 2005). In a more recent study an antisense RNA shown to overaccumulate in the tobacco Nt*rnr1* mutant, was indicated to be the direct cause for the reduction of the 5S rRNA (sense strand) accumulation (Sharwood *et al.* 2009). Hence, as opposed to earlier suggestions, it may well be that transcription of intergenic regions and transcription in antisense orientation to known ORFs is of functional significance in plant mitochondria and not only due to structural peculiarities of the mtDNA or an relaxed control of transcriptional initiation.

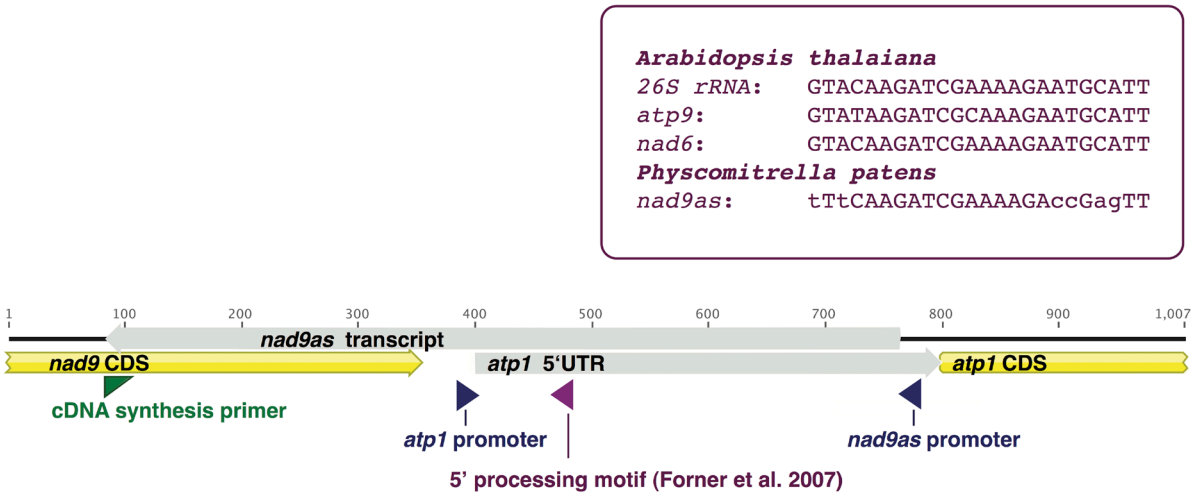


Figure 35: Antisense transcription from a mitochondrial promoter downstream of *nad9*.

4.2 *Physcomitrella* RpoTmp1 and RpoTmp2 are active RNA polymerases

The T7 phage DNA dependent RNA polymerase functions as a single-polypeptide enzyme (Cheetham and Steitz 2000). Thereby it executes all steps of transcription, including promoter recognition, initiation, and elongation, functions that are dependent on and shared by different components in multi-subunit RNA polymerases, such as the bacterial-type polymerase of plastids (PEP). In *Arabidopsis* all three *RpoT* genes were shown to encode transcriptionally active RNAPs. AtRpoTmp displayed no significant promoter specificity, whereas AtRpoTm and AtRpoTp were shown to accurately initiate transcription from overlapping subsets of mitochondrial and plastidial promoters without the aid of protein cofactors (Kühn *et al.* 2007). Since intrinsic promoter specificity was shown to be a feature of both enzymes shared with the T7 RNAP, Kühn et al. suggested, that this trait presumably was conserved in the evolution from an ancestral phage polymerase to the nuclear-encoded mitochondrial and plastidial RNAPs of plants. Because not all promoters mapped *in vivo* were recognised in the *in vitro* system, for such promoters auxiliary factors are likely to be required for an efficient initiation of transcription *in planta*.

Before the third exclusively mitochondrial RNAP, PpRpoTm, was identified in the sequenced genome of *Physcomitrella*, *in vitro* transcription studies aimed to examine the abilities of PpRpoTmp1 and RpoTmp2 to transcribe DNA from mitochondrial promoters located on supercoiled templates, applying the procedure, that was used to study the *Arabidopsis* RpoT proteins (Kühn *et al.* 2007). Differences in the promoter specificity and the ability of both *Physcomitrella* RpoT enzymes to faithfully recognise organellar promoters *in vitro*, were tested with PpRpoTmp1 and PpRpoTmp2 heterologously expressed in *E.coli*.

To omit possible effects of the rather large thioredoxin tag on RNA polymerase activity (Kühn *et al.* 2007), the expression vector pCold I was facilitated, allowing for the expression of His_6-tagged forms of both RpoTmp1 and RpoTmp2. However, the removal of the tag led to instability of both untagged RNA polymerases and therefore prompted His_6-tagged RpoTmp1 and RpoTmp2 to be employed for *in vitro* transcription experiments. The instability of the untagged RpoT proteins was also observed for all three *Arabidopsis* RpoTs (Kühn 2006, Kühn *et al.* 2007) and could result from none-native N-termini (predicted N-terminal transit peptides had been excluded). The most N-terminal amino acid has been shown to influence the protein stability in bacteria and is usually referred to as N-end rule. Therefore, the quality of the N-terminal amino acid while not directly involved in the function of the protein could modulate the turnover of the polypeptide (Meinnel and Giglione 2008).

As reported for all three *Arabidopsis* RpoT enzymes, neither PpRpoTmp1 nor PpRpoTmp2 produced specific RNAs from linear DNA templates. However, unspecific transcriptional rates were markedly different when empty pKL23 linear or supercoiled was used. It has been reported before that the transcriptional activity of purified recombinant protein measured *in vitro*, using calf thymus DNA as a template, is one order of magnitude higher for PpRpoTmp1, than that of PpRpoTmp2 (Kabeya *et al.* 2002). While the preparation of active PpRpoTs might in part influence the enzymatic activity differentially, it is more likely that the absolute activity of PpRpoTmp1 is higher. The difference in absolute transcription rates is reminiscent of the characteristics described for the phage-type RNA polymerases from *Arabidopsis*, where AtRpoTmp exceeds AtRpoTm and AtRpoTp by one order of magnitude as well (Kühn 2006, Kühn *et al.* 2007). Studies of *in vitro* RNA synthesis with untagged *Arabidopsis* RpoTm and RpoTmp did not result in changes of promoter recognition or in changes in the ratio of absolute transcription rates (personal communication A.V. Bohne). Therefore, different RNA synthesis rates and promoter specificities might be related to different functions *in vivo* of RpoTm and RpoTmp as of *Arabidopsis* and RpoTmp1 and RpoTmp2 from *Physcomitrella*.

Table 18 lists promoter sequences included in the study and indicates their utilisation *in vitro*. While specific initiation on supercoiled plasmids by PpRpoTmp1 was confined to the *in vivo* mapped promoter PpP*atp1*-401, PpRpoTmp2 recognised a variety of mitochondrial promoters (Figure 18 and Figure 20). The data suggest that for the promoters shown here to stimulate initiation, sequence specificity is conferred by the RpoT core enzyme without auxiliary factors. Of the promoters tested, diverse sequences displaying CGTA, AATA, CATA and TATA core elements were recognised by RpoTmp2 *in vitro*.

Table 18: *In vitro* Transcription from organellar promoters of *Physcomitrella patens* and *Arabidopsis thaliana*

Promoter	Sequence	PpRpoTmp1	PpRpoTmp2	
Physcomitrella				
P*atp1*-401	CTTGATAATATTCAATTT**CGTA**GTAATTGCATGC	+	+	*mitochondrial*
P*trnE*-549	ATTACTACAAAGAAAAAA**CATA**TATATATATGAA	-	+	
P*atp6*-128	GTATTTCTTCATATATAA**TATA**TATATGAAAAA	-	+	
P*rrn26*-153	CGAATGATGCGATATGTA**AATA**GAAGATACGTAT	-	+	
P*trnfM*-117	CGCCTACGAATGAATCAT**CATA**GAGACTCGCAAT	-	+	
P*trnfM*-36nc	GGAACTGCTGCGCAAAAAGAAATATTTTGCGGCG	-	-	
P*atpB*-270	GTTTT*GAAT*GT**TAAACTAATTA**ATAAAAAGAAAA	-	-	*plastid*
P*atpB*-144	ATAAACATAT**TTATATAATATA**TCAATATAATAA	-	-	
P*atpB*-214	TAAAAAAAAAG**TATAAT**AAACCTAGTTTATAATT	-	-	
P*psbA*-54	ATACATTTTGTGTAA**TACTAT**AAATTAACAAGTT	-	minor	
P*psbA*-81(iv)	ATTTTAAAGATCAGTTGA**CATA**ATAATACATTTT	+	+	
Arabidopsis				*mitochondrial*
AtP*atp6*-1-916/913	TGGTGAAAAGCCCTTTAT**ATTA**TATAAT**A**AAGCG	+	+	
At*atp8*-228/226	TCTTTTTCCATACCATAA**CATA**TAT**AGA**ATCGAT	-	+	
AtP*rrn26*-893	CCGCGGTCCTATCAATTT**CATA**AGA**G**AAGAAAGA	-	-	

Similarly to what is reported for the *Arabidopsis* RpoTs, promoters with an AT-rich sequence following the tetra-nucleotide core element seem to stimulate transcription initiation more efficiently, although PpP*trnfM*-117 gave rise to discernible but not abundant transcripts.

Run-off transcription provided no evidence for transcripts initiated specifically at PpP*atpB*-144 by neither of the two *Physcomitrella* enzymes, but PpP*atp6*-128 was clearly recognised as a promoter indicating that minor deviations from the promoter sequence impede recognition by PpRpoTmp2, or alternatively that sequences up- or downstream from the nucleotides displayed in Table 18 influence promoter strength *in vitro*. Modulated promoter efficiency of T7 RNA polymerase has been documented for upstream sequences extending beyond the promoter boundary (Tang *et al.* 2005).

None of the *atpB* promoters gave rise to specific transcripts. The failure of both PpRpoTmp1 and RpoTmp2 to initiate transcription at PpP*atpB*-144 may be seen as circumstantial evidence that PpP*atpB*-144 is not a NEP promoter (see above). On the other hand only one NEP promoter from *Arabidopsis* has been shown to drive transcription in the minimal *in vitro* system. This has been speculated to be due to the presence of factors activating transcription *in vivo*, but absent in the *Arabidopsis in vitro* transcription system. Considering the diversity of promoters unrecognised *in vitro*, such factors could but do not necessarily need to differ between dissimilar promoter sequences, but might modulate the secondary structure of the DNA allowing for promoter recognition.

Surprisingly, the plastid promoter construct of PpP*psbA*-54 supported specific transcription initiation *in vitro*, by both PpRpoTmp1 and RpoTmp2. Initially PpP*psbA*-54 was included in the assays as a control for a typical PEP promoter containing both conserved -35 and -10 regions, while testing for the ability of P*atpB*-144 to support specific transcriptional initiation.

Transcript sequencing revealed two initiation sites, one minor site with a corresponding 5' end identical to that of transcripts initiated at P*psbA*-54 *in vivo*, allowing to attribute the *in vitro* synthesised RNAs to transcription initiation at a sequence clearly exhibiting all features of a

bona fide PEP promoter element and a predominant TSS 81 nucleotides upstream of the *psbA* translational start site. Re-examination of the *psbA* sequence from P*psbA*-54 revealed the sequence motif TATA to be located 8 nucleotides upstream from the initiating nucleotide. Therefore, the TATA core element as part of the conserved -10 region of the supposed bacterial-type promoter, seems to drive initiation of transcription by both phage-type polymerases *in vitro* (Fig. 19 B and Suppl.Fig.1).

The major signal on the other hand corresponds to a initiating nucleotide that is preceded by a CATA core motif and a very AT-rich upstream region. Interestingly the core CATA motif of the *in vitro* promoter P*psbA*-81 is part of the -35 region of the putative PEP promoter region as mapped *in vivo* (Fig. 19B). Further attempts to identify primary transcripts corresponding to P*psbA*-81 *in vivo* by primer extension and the TAP based method were unsuccessful. Since for both methods only RNA isolated from liquid protonemata cultures was applicable, the importance of the *in vitro* promoter P*psbA*-81 for plastid transcription under different growth conditions and at different developmental stages must remain an open question. Nevertheless, the potential of a bona fide PEP promoter to be driven by RpoT transcription *in vitro* is of interest. If some PEP promoters contain the basic elements for recruitment of RpoT transcription activities this might have provided the fertile grounds for the rise of a NEP activity during the evolution of land plants. Are there examples for such promoters in plastid genomes of flowering plants? Indeed, three cases have been described thus far in *Arabidopsis*, namely AtP*ycf1*-34/39, AtP*rrn16*-112/139 and P*clpP*-115. For initiation on AtP*ycf1*-39 a NEP promoter is suggested to drive transcription and is part of the -10 PEP promoter element, which together with the -35 box likely drives initiation on AtP*ycf1*-34 (Swiatecka-Hagenbruch *et al.* 2007). In the case of AtP*rrn16*-139 the NEP promoter element is suggested to be part of the -35 box of AtP*rrn16*-112, while the -10 element of P*clpP*-115 was suggested to contain a NEP promoter (Swiatecka-Hagenbruch *et al.* 2007).

Thus, if there is no NEP activity in *Physcomitrella* plastids, as suggested by Kabeya and colleges (Kabeya and Sato 2005), PpP*psbA*-54 might be a perfect example of a pre-existing NEP promoter and such promoters might have allowed for the partial takeover of plastid transcription by the former mitochondrial RNAP and resulting in the highly complex plastid transcription machineries of angiosperms.

Clearly, the identification of a third RpoT enzyme in *Physcomitrella* with presumably mitochondrial localisation, calls for PpRpoTm *in vitro* transcription studies. Nevertheless, data presented herein provide evidence that PpRpoTmp1 and PpRpoTmp2 are functional RNA polymerases, which both posses the inherent ability to recognize organellar promoters in a minimal *in vitro* transcription system without the aid of additional cofactors. The difference in the ability of the two tested *Physcomitrella* RpoTs to recognize promoter structures is reminiscent of the characteristics described for the RpoT proteins from *Arabidopsis*, where AtRpoTmp displays no significant promoter specificity, while AtRpoTm and AtRpoTp accurately initiate transcription from overlapping subsets of mitochondrial and plastidial promoters. The two *Physcomitrella* polypeptides were shown to differ in unspecific transcriptional activity before, and the possibility was put forward that PpRpoTmp2 due to its low activity could be dispensable (Kabeya *et al.* 2002). As for *Physcomitrella* RpoTmp2 the lower unspecific transcription activity of *Arabidopsis* RpoTm is correlated, with the ability to recognize numerous mitochondrial promoters *in vitro*. While AtRpoTm was therefore suggested to function as the basic RNA polymerase in mitochondria, the inability to isolate plants homozygous for T-DNA insertions in *AtRpoTm* provides direct evidence that the purely mitochondrial RNAP is unconditionally required for the transcription of most mitochondrial

genes in *Arabidopsis*. To the contrary, *Physcomitrella* RpoTmp2 knock out is not lethal and therefore indicates that its function can be partially substituted by one or both of the other RpoT proteins. Phenotypical aberrations of *rpoTmp2* plants however suggest that RpoTmp2 is important for normal growth and development.

4.2.1 *Physcomitrella* enzymes recognize mitochondrial promoters from *Arabidopsis in vitro*

While the promoter AtP*rrn26*-893, only weakly recognised by *Arabidopsis* RpoTm and RpoTp, did not result in recognisable promoter recognition *in vitro* by the two *Physcomitrella* enzymes, data provided for the mitochondrial promoter P*atp8*-228/226 of *Arabidopsis* are supportive of cross-species promoter recognition by both *Physcomitrella* phage-type RNA polymerases. However, recognition of P*atp6*-1-916/913 by RpoTmp2, but not RpoTmp1 is indicative of differential properties of both enzymes (Tab. 18). The last common ancestor of both eudicots and mosses lived about 450 million years ago, while the divergence time of the two *Arabidopsis* RpoT enzymes RpoTm and RpoTmp as found in mitochondria is only a minimal fraction of that (see 3.3.2). Moreover gene duplication events giving rise to multiple RpoTs in both *Physcomitrella* and *Arabidopsis* are clearly independent events (Richter *et al.* 2002). Therefore, data presented herein might suggest that coding for two RpoT proteins one representing an enzyme with a high portion of non-specific transcriptional activity, as seen for AtRpoTmp and PpRpoTmp1 and one that can act as a single-polypeptide enzyme and recognize numerous mitochondrial promoters *in vitro* as AtRpoTm and PpRpoTmp2 echo convergent inventions but reflect complementing roles of these RNA polymerases in plant mitochondrial transcription.

4.3 The phenotype of PprpoTmp is reminiscent of mitochondrial mutants

When this study was started very little was known about the potential division of labour between the RpoT proteins in organellar transcription. *RpoTm* and *RpoTmp* genes in *Arabidopsis* had been reported to display overlapping expression patterns (Emanuel *et al.* 2006). Analysis of an RpoTmp null mutant revealed that the lack of RpoTmp resulted in induction of several plastid genes in dark-grown seedlings upon illumination (Baba *et al.* 2004), suggesting RpoTmp to be the key enzyme transcribing plastid and mitochondrial genes during early seedling development, while RpoTm and RpoTp would take over transcription in the respective organelle later in development.

To elucidate potentially distinct functions of RpoTmp1 and RpoTmp2 corresponding genes were targeted for gene knock-out by homologous recombination. The identification of six independent *RpoTmp2* knock-out lines is implicit of a dispensable function in the moss or alternatively, a function of RpoTmp2 that can be complemented at least in part by another RNA polymerase. In contrast no *PprpoTmp1* homoplasmic moss lines could be discovered and transformants were found to be 2n possessing both the wild-type and the *PprpoTmp1-nptII* allele. This in turn suggests that *PpRpoTmp1* has an indispensable function in *Physcomitrella patens* and disruption of the *PpRpoTmp1* locus is likely lethal.

While no changes in the morphology of either organelle were observable by EM, the aberrations in leaflet morphology displayed by *PprpoTmp2* plants are reminiscent of leaf

phenotypes described for mutants impaired in mitochondrial function (de Longevialle *et al.* 2007, Reichheld *et al.* 2007, Van Aken *et al.* 2007). The developmental delay throughout all stages of the life-cycle, observed in rpoTmp mutants is reported to be a common feature of mitochondrial mutants (Zsigmond *et al.* 2008). More importantly analysis of a developmentally and growth retarded *Physcomitrella* RecA1 mutant revealed mitochondrial genome stability to be disrupted (Odahara *et al.* 2009). The overlapping phenotype of recA1 and *rpoTmp2* plants as shown in Figure 36, therefore might suggest that the phenotypical aberrations observed in *rpoTmp2* plants are possibly caused by a mitochondrial dysfunction due to the lack of functional PpRpoTmp2.

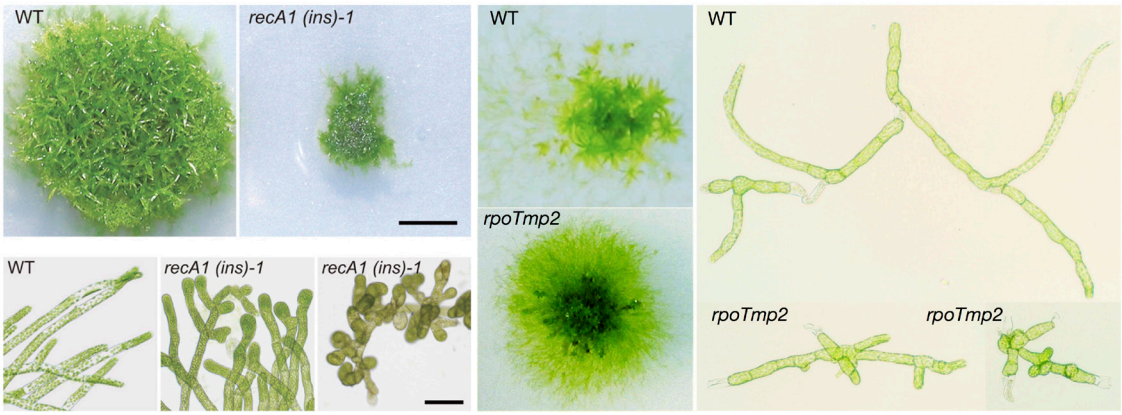


Figure 36: Phenotypes of RecA1 and *PprpoTmp2* mutants.

Mitochondrial dysfunction has been reported to have various effects on plant growth and development. As shown for complex I deficiencies plants can not sustain a high rate of growth and some cells may even prove unviable during specific points in the life cycle, such as seed set, germination, tissue formation and leaf growth (Garmier *et al.* 2008).

In plants, one major target of the retrograde response is the induction of non-proton-pumping NADH dehydrogenases and the alternative oxidase (AOX), both of which provide substitute pathways for the maintenance of redox balance and mitochondrial metabolism under conditions when OXPHOS is inhibited. Such circumstances occur naturally in photosynthetic plants, since in the light, photosynthesis produces ATP at high levels. This in turn is a limiting factor for mitochondrial OXPHOS due to coupling. Therefore, expression of the substitute pathways allows for NADH reoxidation and in turn operation of the TCA cycle that produces intermediates of several biosynthetic pathways (Fernie *et al.* 2004, Noguchi and Yoshida 2008, Rasmusson *et al.* 2008). Studies on AOX have revealed that its expression is induced by a variety of treatments often referred to as stresses, such as the use of inhibitors of the mitochondrial electron transport chain (Clifton *et al.* 2005, Van Aken *et al.* 2009) or in plants with mutations of respiratory complex subunits (Dutilleul *et al.* 2003, Karpova *et al.* 2002, Noctor *et al.* 2004). Additionally induction of AOX is commonly used as a marker for the mitochondrial retrograde response in plants beside its function in the plant mitochondrial electron transport chain. In *PprpoTmp2* plants AOX protein levels were elevated, suggesting an imbalance in the respiratory chain function. On the other hand recently it has been shown that excess light in a mutant lacking functional FtsH2, which results in plastid photo-oxidative stress, can induce changes in AOX expression (Yoshida *et al.* 2008). Nevertheless this leaf-

variegated yellow variegated 2 mutant has a very strong PSII phenotype, not observed in *PprpoTmp2* plants. Taken together, this suggests, that elevated AOX levels in *PprpoTmp2* are more likely due to an imbalance in mitochondrial function.

4.3.1 Transcript accumulation is imbalanced in PprpoTmp

The accumulation of numerous mitochondrial transcripts was found to be enhanced in the absence of PpRpoTmp2 as shown by dot-blot analysis and qRT-PCR, possibly due to the enhanced expression of PpRpoTmp1 and PpRpoTm, indicated by elevated transcript levels of PpRpoTmp1 and PpRpoTm. While this finding suggests the existence of a regulatory mechanism to compensate for the loss of RpoTmp2 function in the mutant, the observable elevation of mitochondrial transcripts could, while rescuing transcription of functionally important genes like atp6 to wild-type levels, result in the over accumulation of deleterious transcripts. As described below, such an imbalance in the transcript pool could possibly effect posttranscriptional processes such as transcript processing, editing and translation and be causative for the phenotype, that was observed in *PprpoTmp2* plants. While all promoters identified in the wild type were also found to be active in *PprpoTmp2* and therefore the data presented in this work do not provide evidence for a promoter specific function of PpRpoTmp2, the observation of additional transcription start sites in *PprpoTmp2* plants (P*trnfM*-117, P*trnfM*-36) might be an indicator of an imbalanced mitochondrial transcription.

Several authors have suggested that antisense and intergenic transcripts have no functional significance, but are instead a disadvantageous result of the relaxed transcriptional control in *Arabidopsis* mitochondria (Forner *et al.* 2007, Holec *et al.* 2006, Kühn 2006, Kühn *et al.* 2005, Perrin *et al.* 2004).

In plant mitochondria such RNAs described as illegitimate transcripts are shown to be polyadenylated. The poly(A) tails are implicated in triggering the degradation of such transcripts through the 3'-5' exonucleolytic activity of polynucleotide phosphorylase, suggesting that illegitimate transcript synthesis is compensated by RNA degradation mechanisms (Holec *et al.* 2008a). Plants with reduced PNPase levels have been shown to exhibit developmental defects with pronounced aberrations in leaf shape (Perrin *et al.* 2004). These defects have been speculated to be at least in part due to the accumulation of deleterious transcripts (Holec *et al.* 2008a).

In *S. cerevisiae* mitochondria SUV3 is the master 3’ to 5’ exoribonuclease responsible for RNA turnover and surveillance. Inactivation of SUV3 leads to the disruption of mitochondrial gene expression and loss of respiratory function, likely caused by over-accumulation of aberrant transcripts and precursors. Interestingly, one of the suppressor loci that partially restored mitochondrial function in suv3Δ strains was found to correspond to partial loss-of-function mutations in the *Rpo41* gene, which encodes the mitochondrial RNA polymerase (Rogowska *et al.* 2006). This directly indicates, that reduced transcription rescues defects in RNA turnover and thereby demonstrates the vital importance of maintaining a balance between RNA synthesis and degradation. Indeed cytoplasmic male sterile (CMS) hybrid lines have been shown to have elevated mitochondrial transcript levels, while no differences were found in the mitochondrial genome structure (Yamasaki *et al.* 2004). The alloplasmic hybrid of *Triticum aestivum* with a cytoplasm of wheatgrass (*Agropyron trichophorum*) shows highly depressed vigor and complete male sterility (Suzuki *et al.* 1995). While mitochondrial

transcript levels of 10 different genes were strongly elevated in these CMS plants, differences in the mitochondrial genome structure were not observed between the depressed and a restored line. In the CMS line the wild-type phenotype and male fertility was shown to be restored by the presence of one short-arm telocentric homeologous group 1 chromosome (telosome) from the cytoplasm donor, and paralleled by restoration of 'normal' mitochondrial transcriptional activity (Suzuki *et al.* 1995). This suggests that the observed elevation of mitochondrial transcripts, due to a higher transcriptional rate results in depressed vigor and male sterility of the alloplasmic hybrid.

High abundance of the whole mitochondrial transcriptome, including cryptic transcripts may exert adverse effects through depletion of factors involved in post transcriptional processes, perturbation of gene expression through antisense RNA, imbalance the ratio of respiratory complex subunits or result in the production of “toxic” proteins such as in some CMS mutants.

Abundant illegitimate transcripts have been shown to affect the editing rate of endogenous sites in plastids likely by depleting editing factors (Chateigner-Boutin and Hanson 2002, Hayes *et al.* 2006), indicating that some elements of the editing machinery are limited and that transcripts compete for such components in chloroplasts. Due to the similarities of the editing processes in plastids and mitochondria (Hiesel *et al.* 1989, Pring *et al.* 1993), such a scenario might apply to mitochondria as well. Moreover, abundant cryptic transcripts could also compete for factors involved in the processing of 3′ and 5′ ends in a similar fashion, though data supporting such a scenario are lacking to date.

As discussed above antisense RNA could potentially affect post-transcriptional processes in mitochondria, though data are lacking to support such a conclusion. However, in chloroplasts it was shown that expression of short antisense RNA in trans, though not affecting the sense transcript, resulted in a strong morphological phenotype for as yet unknown reasons (Hegeman *et al.* 2005). This in turn illustrates, that aberrant antisense RNAs might have harmful effects on plant organelles.

Another level of mitochondrial gene expression that could be negatively impacted by disturbances of the transcript pool is translation. The synthesis of respiratory complexes is dependent on the coordinated expression of subunits encoded in the nucleus and on the mitochondrial genome. Stress induction by sucrose starvation has been shown to result in a marked elevation of mitochondrial transcripts. But this elevation did not affect translational performance as measured by an *in organello* translation assay or steady state protein levels (Giegé *et al.* 2005). Giege and colleagues therefore suggested that mitochondrial biogenesis is regulated by changes in nuclear gene expression and coordinated at the posttranslational level, i.e., through complex assembly. In concordance with these data, elevated transcripts in *PprpoTmp2* plants were not represented by steady state protein levels of atp1 and nad9, though a slight increase in the abundance of cox2 was observable.

Whether the mitochondrial dysfunction in *rpoTmp2* plants is directly due to the scarcity of a specific transcript as a consequence of lacking RpoTmp2 transcriptional activity, though no significant reduction was found among 32 mitochondrial transcripts tested, or due to a imbalance in the mitochondrial transcript pool supported by the observation of a general elevation in mitochondrial steady state transcript levels in *PprpoTmp2* plants remains open.

Advanced whole transcriptome deep sequencing analysis for the mitochondrial genome would allow to address this question, since it not only provides data for the known orfs, but also provides high resolution signals for intergenic regions and for both strands of the investigated genome.

4.4 Expression of *RpoT* genes in the life cycle of *P.patens* and the matter of dual targeting

The self-fertile moss *P. patens* is a simple plant whose primary life stage consists of filamentous protonema cell stage, with a succession of chloronema and caulonema cells and small leafy shoots (gametophores), all of which are haploid. On the apices of the gametophores both sex organs archegonia (female) and antheridia (male) are formed. After fertilisation the diploid mainly heterotrophic sporophyte grows from the zygote and is supported by the photoautotrophic gametophyte. Nothing is known about a potential regulation of plastid or mitochondrial transcription throughout these developmental stages. Quantitative real time RT-PCR analysis was facilitated to provide evidence for changes in the expression profile of all three *RpoT* genes throughout the *Physcomitrella* life-cycle.

Protonematal cells contain large mitochondria and developing chloroplasts. In this respect, protonemata were suggested to be most homologous to the developing seedlings of flowering plants. Since RpoT proteins in flowering plants have been shown to be active in young developing tissues, this concept and the ease to grow protonema in liquid culture in sufficient amounts prompted experiments in this study to be based on this part of the life-cycle. When *RpoT* transcript steady-state levels were compared between different stages of the *Physcomitrella* life-cycle, normalised with the respective *PpRpoT* transcript levels measured in gametophytes, the expression of *PpRpoTmp1* and *PpRpoTm* genes surprisingly was found to be induced in sexual gametophytes, but reduced at the protonema stage.

PpRpoTmp2 appeared to be stably expressed throughout the *Physcomitrella* life-cycle. Since the RNA prepared from sexual gametophytes contains RNA pooled from late gametophytes, archegonia, antheridia and sporophytes it has to be assumed, that *PpRpoTmp1* and *PpRpoTm* gene expression is elevated at least during one of these stages. Since in *Arabidopsis* the AtRpoTp enzyme exclusively found in plastids plays a major role in chloroplast transcription, while the plastidial portion of AtRpoTmp likely has its main function in the transcription of genes in non-green tissues (Azevedo *et al.* 2008, Courtois *et al.* 2007, Emanuel *et al.* 2006, Swiatecka-Hagenbruch *et al.* 2008), it is tempting to speculate that elevated gene expression of *Physcomitrella* RpoTm and RpoTmp1 might be correlated with a major role of both proteins in the organellar transcription of the mainly heterotrophic sporophyte. The sporophyte might depend on elevated mitochondrial activity, or more specifically on OXPHOS through respiratory chain function, due to the absence of photosynthetic ATP production. Mainly due to the biochemical inaccessibility of the sporophyte, no evidence has been provided for the presence or absence of PpRpoTmp1 and PpRpoTmp2 in sporophytic plastids. Nevertheless dual targeting properties of both PpRpoTmp1 and PpRpoTmp2 signal peptides are suggestive of a function in plastids of *Physcomitrella patens* (Richter *et al.* 2002). Since targeting studies revealed that both enzymes are likely not found in plastids in neither protonema nor gametophytes (Kabeya and Sato 2005), conservation of the plastid targeting peptides indicates their functioning in other developmental stages of the moss or under conditions not tested in aforementioned assays.

The retention of the N-terminal extension that encodes a plastid targeting sequence in PpRpoTmp1 and PpRpoTmp2 must be considered meaningful. To the contrary Kabeya and colleagues argue that since sequence and length are not conserved between the transit peptides, this would in turn indicate that there is no selective pressure on the N-terminal extensions (Kabeya and Sato 2005). Their conclusion that the two PpRpoTs are targeted only to mitochondria due to exclusive translation from the second AUG codon in every cell and developmental stage must be taken cautiously. If the two transit peptides are homologous, the preservation of their ability to promote plastid targeting, albeit having diverged considerably, as indicated by the alteration of their primary structure, is rather strong evidence for selective pressure on both N-terminal extensions. In the reported targeting experiments, the over expression of GFP-fusion constructs might obstruct possible regulatory mechanisms that could influence translational initiation. Since initiation of translation from the first AUG was repressed for all constructs with the native 5'UTR in their study, this could be seen as circumstantial evidence that translation initiation on both *RpoTmp1* and *RpoTmp2* transcripts is negatively regulated in *cis* and could be positively affected in *trans in planta* (Kabeya and Sato 2005). Nevertheless, results of their immunological analyses provide important inside on the apparent lack of plastidial RpoTmp1 and RpoTmp2 in protonemata of *P. patens* (Kabeya and Sato 2005). Both enzymes have been shown to encode purely mitochondrial transit peptides when translation is initiated from the second start codon of the coding frame (Kabeya *et al.* 2002, Kabeya and Sato 2005, Richter *et al.* 2002). Therefore, the identification of a third phage-type RNAP coding frame shown to be expressed through all stages of the *Physcomitrella* life-cycle (Fig. 21 and 24) and likely exclusively localised in mitochondria (Fig. 23) was rather surprising. Considering the divergence of all three RpoT proteins in their primary structure (see 3.1.6), conservation and selection for their retention has to be assumed. As judged from the branch length in the phylogenetic tree, the two duplication events of the three *Physcomitrella* RpoTs, though lineage specific, are rather not species specific but considerably older. This observation gains weight to the functional importance of the N-terminal extensions encoding plastid targeting properties of PpRpoTmp1 and PpRpoTmp2, but also argues for a significant function of all three proteins in the *Physcomitrella* lineage.

The apparent complexity of the mitochondrial transcription apparatus with three divergent phage-type polymerases and the potentially even more complex plastid transcription system with the dedicated bacterial-type multi-subunit polymerase with two nuclear encoded RpoA subunits, a multiplicity of sigma factors and potentially two additional phage-type proteins, contrasts starkly with the reduction of these genomes during the course of the integration of the endosymbiont into the eukaryotic cell. High complexity in organellar RNA metabolism seems to be a common trend in the evolution of land plants. One conceivable explanation for this tendency is based on the assumption that complexity of organellar RNA metabolism might be the direct result of an increased demand for regulating steps in organellar gene expression, possibly to integrate the organelle into the context of varying demands in plant cells after the transition to land. In this context multiplicity and differential expression of RpoTs as observed in *Physcomitrella* as well as in *Arabidopsis* would suggest an elevated importance for the regulation of organellar transcription in land plants.

Alternatively, Maier and colleagues have put forward another hypothesis to explain the come about of highly complex RNA metabolism specifically for the evolution of the plastid genome (Maier *et al.* 2008), which due to the observed highly complex RNA metabolism in land plant mitochondria might be easily extended to this organelle. Briefly, this hypothesis proposes the expansion of complexity in organellar gene expression to be selectively neutral, driven by varying mutation rates and differences in the strength of genetic drift between the organellar

and nuclear genomes of plants. In an *Arabidopsis* Sig6 knockout mutant an upstream-located NEP promoter is suggested to compensate for lack of transcription from the main *atpB* PEP promoter (Schweer *et al.* 2006). The authors therefore proposed a function for NEP as an SOS RNAP activity in plastid transcription. In accordance with the 'drift hypothesis' the nuclear encoded NEP initially might have represented an RNAP activity able to compensate for mutations in original PEP promoter elements or in components important for the expression of the plastid encoded bacterial-type polymerase. Since AtRpoTp and AtRpoTmp as well as PpRpoTmp1 and PpRpoTmp2 likely represent RNAPs with very different characteristics, as indicated by the *in vitro* transcription experiments, they could provide specific functional complements for the plant organelle transcription apparatus in such a scenario. The varying expression of all *Physcomitrella RpoTs* throughout the life cycle of the moss might in turn be interpreted as a subsequent specialisation.

4.5 Analysis of deviating functions of two phage-type RNA polymerases in mitochondria of *Arabidopsis thaliana*

Mitochondrial transcription in higher plants has been found to be non-stringently controlled (Holec *et al.* 2008a), with transcription starting from multiple promoters not restricted to the upstream regions of functionally important coding regions. The transcriptional machinery thereby continuously generates non-functional or even detrimental RNAs. These transcripts have been shown to be efficiently removed by dedicated RNA-degrading enzymes in mitochondria (Holec *et al.* 2006). Moreover such post-transcriptional processes are also suggested to be the major regulator of functionally important mitochondrial RNAs. This suggestion is based on the observations that for some genes transcription rates are decoupled from the abundances of steady state transcript levels and that transcription rates of genes encoding components of the same multi-subunit complex can differ significantly (Giegé *et al.* 2000). Though tissue- and cell-specific variation in mitochondrial gene expression has been correlated with transcriptional levels by some authors (Li *et al.* 1996, Topping and Leaver 1990) mitochondrial transcription has been lately predicted to play only a minor or no part in the regulation of mitochondrial function (Giegé *et al.* 2005). Comprehensive mapping of multiple mitochondrial TSS did not support a role of distinct promoters in developmental or tissue-specific regulation of mitochondrial gene expression in *Arabidopsis* and was therefore concluded to be in favour of the established paradigm (Kühn *et al.* 2005).

The *AtRpoTm* and *AtRpoTmp* genes in *Arabidopsis* have been reported to display overlapping expression patterns in different tissues and at different developmental stages (Emanuel *et al.* 2006) and were therefore proposed to transcribe at least partially different sets of mitochondrial genes. In a study of an *Arabidopsis* line lacking AtRpoTmp (Baba *et al.* 2004) both AtRpoTm and AtRpoTp have been suggested to be important at a later developmental stage, while AtRpoTmp was proposed to be the key RNA polymerase transcribing organellar genes during early seedling development. In plastids AtRpoTmp and AtRpoTp have been suggested to be partially redundant since insertion lines with both enzymes disrupted were seedling lethal (Hricova *et al.* 2006) but plants lacking either AtRpoTmp or AtRpoTp showed similar phenotypical alterations when compared to the wild type.

The plastid *rrn16* PC promoter has been shown to be highly RpoTmp-specific and nearly inactive in the absence of RpoTmp in *Arabidopsis* (Courtois *et al.* 2007). The developmentally restricted role and the inactivation of the PC promoter following germination has been reported to be due to the immobilisation of plastidial RpoTmp at the thylakoid membrane (Azevedo *et al.* 2008). This sequestration of the enzyme to the membrane is reportedly not found for the mitochondrial portion of RpoTmp (Azevedo *et al.* 2008). In addition to functioning at the Pc promoter, RpoTmp is clearly capable of initiating transcription from several other plastidial promoters (Swiatecka-Hagenbruch *et al.* 2008).

However, plants lacking AtRpoTmp have been shown to display no changes in the accumulation of several mitochondrial transcripts tested (Baba *et al.* 2004). Moreover AtRpoTm, but not AtRpoTmp, specifically recognised mitochondrial promoters *in vitro* (Kühn *et al.* 2007), indicating a significant role for the purely mitochondrial enzyme in mitochondrial gene expression. Therefore, the transcriptional role of AtRpoTmp in mitochondria and a possible division of labour between the two mitochondrial RNA polymerases was still unclear.

4.5.1 AtRpoTm disruption is lethal

Neither in our group nor in the laboratory of Prof. Whelan plants homozygous for a T-DNA insertion in *AtRpoTm* could be isolated. Transmission of *AtrpoTm* alleles by both male and female gametes were found to be reduced by two-thirds in reciprocally backcrossed plants in the laboratory of Prof. Whelan, indicating decreased *AtrpoTm* gamete fitness (Kühn *et al.* 2009). These findings have recently been confirmed by the comprehensive characterization of two rpoTm mutant alleles mgp 3-1 (male gametophyte- defective 3-1) and mgp 3-2 (male gametophyte- defective 3-2) identified in a screen for male gametophyte-defective mutants in *Arabidopsis thaliana* (Tan *et al.* 2010).

The mgp3 mutations caused embryo lethality and partially impaired female gametophyte development. While pollen formation and pollen germination were reported to be normal, pollen tube growth was found to be significantly retarded. Taken together, the observations suggest that disruption of AtRpoTm is lethal and indicate AtRpoTm to be the major transcriptional activity in *Arabidopsis* mitochondria. In contrast homozygosity of T-DNA insertions in the *AtRpoTmp* locus was not lethal, but produced a phenotype characterised by a developmental delay and wrinkly rosette leaves (Baba *et al.* 2004, Courtois *et al.* 2007).

As described before the rosette leaf morphology displayed by *AtrpoTmp* plants is reminiscent of the leaf phenotype described for mutants impaired in mitochondrial function (de Longevialle *et al.* 2007, Reichheld *et al.* 2007, Van Aken *et al.* 2007) and similar aberrations in leaflet morphology were observed for *PprpoTmp2* plants (see 3.1.2). These observations prompted the analysis of mitochondrial transcript accumulation in *AtrpoTmp* seedlings and rosette leaves.

4.5.2 RpoTmp is required for the transcription of distinct mitochondrial genes

Surprisingly, strong changes in the abundances of numerous mitochondrial transcripts in three mutant lines were found consistently for rpoTmp-1, rpoTmp-2, and rpoTmp-3 when compared to the wild type (Fig. 26). Since Baba and colleagues reported mitochondrial transcript abundances not to be affected in rpoTmp plants (Baba *et al.* 2004), the qRT-PCR results were validated through RNA gel blots using strand specific RNA probes. Deviations in the level but not the trend of the observed changes between qRT-PCR and northern results are likely due to the fact that qRT-PCR cannot distinguish sense from antisense transcripts. Different growth conditions and/or different ages of the seedlings used in the analysis might have prohibited the observation of changes in the steady-state transcript pools by Baba and colleagues. The alterations of the steady state transcript profiles observed in *rpoTmp* mutant plants found to be stable from 7 day old seedlings to mature plants grown for 9 weeks on soil are in line with the reported stable expression and role of AtRpoTmp throughout the whole *Arabidopsis* life cycle specifically in tissues with elevated mitochondrial activity (Emanuel *et al.* 2006).

The run-on transcription assays revealed transcriptional changes in *AtrpoTmp* to coincide with the measured differences in transcript accumulation. Although reduction in transcription of *rps4* and *nad2*e1_e2 in the mutant was not as severe as the measured decrease in transcript accumulation, implicating a contribution of altered transcript stabilities to changed transcript abundances in *AtrpoTmp*, transcriptional changes clearly account for the changes in transcript abundances in the mutant. Therefore, the severe reduction in mitochondrial transcription in *AtrpoTmp* plants establishes the mitochondrial portion of AtRpoTmp to serve a significant role in the synthesis of mitochondrial RNAs. Since some mitochondrial genes were transcribed to reduced levels, while others were unaltered or elevated, AtRpoTmp function in mitochondria seems to be gene specific.

Surprisingly, post transcriptional processes such as stabilisation of RNAs, indicated to be the main modifier of an supposed relaxed transcriptional machinery in plant mitochondria or enhanced translation are not showing balancing effects but rather the opposite was observed as seen for *rps4*. This suggests that regulatory post transcriptional processes are not acting on the transcripts specifically reduced in *AtrpoTmp* and/or can not counterbalance the severe reduction of transcription in *AtrpoTmp* plants demonstrated by the reduced levels of respiratory chain complexes I and IV in *rpoTmp* mutants. The reduction in the steady state levels of both respiratory complexes are likely the direct result of diminished amounts of the *nad6* and *cox1* transcripts. The reduction of *rps4* has no detectable effect on mitochondrial translation, since expression of all other complexes is unaffected and *in organello* protein synthesis assays found no evidence for inhibition of translation in mutant mitochondria.

The inability to compensate for the reduction of transcription might also indicate a lack of feedback regulation for a specific set of transcripts in plant mitochondria. However, elevated mitochondrial DNA copy number per nuclear genome as observed in *AtrpoTmp* plants might be a SOS response to disturbances in the gene expression machinery or more general to energy constraints caused by defects in the energy-converting organelles by increasing cellular mtDNA levels. In line with this hypothesis a barley mutant with impaired chloroplast function shows elevated mitochondrial gene copy number (Hedtke *et al.* 1999b).

In the aforementioned barley mutant the increase in mitochondrial gene copy number was also accompanied by elevated levels of mitochondrial transcripts. Interestingly, elevation of

transcriptional performance and transcript accumulation observed for numerous mitochondrial transcripts in *AtrpoTmp* plants did not exceed the increase in gene copy number per cell. Therefore, enhanced transcription and transcript accumulation is likely due to elevated cellular mtDNA levels in mutant plants and in line with the observation that AtRpoTm expression is only slightly enhanced in leaves of *AtrpoTmp* plants (Baba *et al.* 2004). However, Preuten et al. only recently showed that gene copy numbers and transcript levels are not correlated during leaf development in wild-type *Arabidopsis* plants (Preuten *et al.* 2010). Therefore, a contribution of elevated RpoTm activity possibly together with the elevation of gene copy numbers in *AtrpoTmp* plants cannot be dismissed completely to account for the observed elevation in a subset of the transcript pool. This contrasts to some extent with the observations made in *Physcomitrella rpoTmp2* plants, where the gene copy number per cell was not increased when compared to the wild type. Instead *PpRpoTmp1* expression was elevated 4-6 fold and might therefore account for the elevation of numerous mitochondrial transcripts, indicating that different mechanisms in response to defects in the energy-converting organelles might be employed in both plants.

If elevation of transcription in *AtrpoTmp* plants is indeed due to the higher copy number, the extent of the reduction in RNA synthesis per gene copy for the negatively affected genes would be even more striking. Accordingly, transcripts with unchanged or only mildly elevated abundance would be still made at lower levels per gene. While transcription of *ccmFc* is likely independent of AtRpoTmp, this in turn would expand the role of AtRpoTmp in the transcription at least in part to genes such as *rps3* and *atp6-1*. The synthesis of those RNAs could either be mainly carried out by RpoTm and depend on RpoTmp to a minor extent or alternatively genes such as *rps3* are mainly transcribed by RpoTmp in wild-type plants, but RpoTm is able to substitute in mutants lacking RpoTmp.

4.5.3 RpoTmp activity is gene-specific rather than promoter-specific

Surprisingly, the data provided by the extensive analysis of TSS in *AtrpoTmp* plants did not indicate lacking transcriptional initiation from a specific subset of promoters preceding the genes for which the reduction of transcript accumulation had been observed. To the contrary all promoters identified in the wild type were also found to be active in mutant plants. As a reminder it should be noted here that as in *AtrpoTmp* plants no changes in the utilisation of organellar promoters had been observed between *Physcomitrella* wild-type and *PprpoTmp2* plants (see 3.1.4). In the comprehensive promoter analysis for 22 out of 30 promoters characterised in this study, evidence provided by the 5'-RACE technique was sufficient to validate them as transcription start sites. The remaining eight 5' ends (see Tab. 17) lacked distinguishable 5’ triphosphates but promoter-like sequences were found directly upstream. The poor stability of triphosphates at several known primary transcript 5' ends in mitochondria has been reported before (Kühn *et al.* 2005). Taken together, these observations suggest those termini being primary rather than processed.

As in mitochondria of *Physcomitrella rpoTmp2* plants where an additional promoter was observed (see Fig. 14) the promoter P*ccmFc*-743 is found consistently only in *AtrpoTmp* plants. Primary 5' ends mapping to P*ccmFc*-743 were consistently detected in mutants but not in the wild type, indicating that P*ccmFc*-743 is inactive in the wild type. Notably, especially

high levels of the *ccmFc* transcript might therefore in part due to the additional use of P*ccmFc*-743. Intuitively a promoter used in the absence of RpoTmp should still be used when the enzyme is active. Without assuming lack of transcriptional initiation at P*ccmFc*-743 in the wild type, rapid processing of the corresponding 5'transcript end in wild-type plants could also account for this observation. The latter explanation would in turn argue for differences in the post transcriptional processing between mutant and wild-type plants. This question was not specifically addressed for all mitochondrial transcripts in the work presented here and must therefore remain open.

Nevertheless, lacking initiation from certain mitochondrial promoters was not the reason for the observed reduction of transcript accumulation in mitochondria of plants lacking AtpoTmp. Therefore, it was speculated that AtRpoTm might be able to substitute for AtRpoTmp at promoters of a certain architecture but with lesser efficiency. This prompted the analysis of 5' end abundance in mutants and wild type by quantitative ribonuclease protection assays (RPAs). RPAs were performed on 5' termini of transcripts reduced in abundance in rpoTmp (*nad1*e4 and *matR, nad2*e1_e2, *nad6*, *cox1*, and *ccmC*) as well as on a transcript accumulating more abundantly in mutants (*atp6*-1). These assays revealed a surprising result. All analyzed 5' termini, primary or processed, were reduced in abundance in rpoTmp but termini of *atp6-1*. Signals corresponding to P*atp6*-1-156 and P*atp6*-1- 200 were in fact enhanced in mutants. Therefore, quantitative analysis of transcript 5' ends clearly showed all 5' termini of a transcript with reduced transcript accumulation to be diminished in *AtrpoTmp* plants. Since lowered abundances of subgenomic DNA molecules in *AtrpoTmp* plants were not observed, promoter mapping together with RPA and run-on transcription data indicate that reduced transcriptional initiation on multiple and diverse promoters, preceding genes with lowered transcript accumulation, are likely the direct cause for the reduction of respiratory complexes I and IV in the mutant.

Since transcription in plant mitochondria is not efficiently terminated at dedicated structures downstream of mitochondrial ORFs, co-transcription of closely spaced coding regions could account for the reduced transcript abbundance of multiple genes. For the *rps4* locus promoter mapping and RPA data show that even though *rps4* and *nad2*e1_e2 appear to be partly co-transcribed, transcription from promoters mapped in the intergenic region of both coding frames is also reduced. Furthermore *nad2*e1_e2 and nad6 are likely not transcriptionally linked and the two *nad6* specific promoters downstream from *nad2*e1_e2 also exhibit reduced transcription initiation. Similar observations were made for the *nad1*e4-*matR*-*nad1*e5 locus. While *matR* seems to be exclusively cotranscribed with the upstream *nad1*e4, the promoters in the intergenic region of *matR* and *nad1*e5 confer reduced transcription. Therefore, only the reduction of *matR* transcripts is exclusively attributable to the co-transcription with an upstream coding region, while for all other genes, even if closely spaced, reduction of one dominant polycistronic message is unlikely to account for their reduced transcript levels. Instead all promoters mapped in the described loci were less actively used in *AtrpoTmp* plants.

A comparison of promoters from which transcription initiation was shown to be reduced in *AtrpoTmp*, and all other known mitochondrial promoters in *Arabidopsis* did not detect any common motif, that would identify them as a distinct subset of mitochondrial promoters. The only sequence element found to be statistically overrepresented amongst promoters conferring

reduced transcriptional activity in the mutant is the GAAA upstream tetranucleotide, also found to precede mitochondrial promoters mapped in *Physcomitrella patens*.

Since not all of the promoters of genes transcribed less actively in the *Arabidopsis* rpoTmp mutant contain this element and the positioning is rather variable (Fig. 30), this element alone is unlikely to confer RpoTmp-specific promoter recognition or repression of the ability of AtRpoTm to effectively initiate transcription when substituting for RpoTmp in the mutant.

Of interest none of the promoters preceding genes with reduced transcription contained the motifs CATAGAGAA nor CGTATATAA, associating those motifs specifically with AtRpoTm function. Nevertheless, several of the promoters unremarkably used by RpoTm in mutant plants contain extensively divergent core motifs. Therefore, sequences such as CATAGAGAA seem to represent only a small portion of promoters employed by AtRpoTm.

Thus the only criterion to define a distinct subset of mitochondrial promoters was reduced transcription initiation in *AtrpoTmp* plants. As a result control of mitochondrial transcription in the eudicot plant *Arabidopsis thaliana* might be best described as partially conferred through elements within the transcribed regions rather than by sequence elements upstream of the TSS, possibly involving RpoTmp-specific *trans* factors (Fig. 37A). In such a model mitochondrial promoters of genes transcribed mainly by AtRpoTmp may simply function as DNA elements that permit easy melting of the double helix to initiate transcription. In support of such a model mitochondrial promoter selection has been reported to be under the controll of nucleus-encoded factors in alloplasmic lines of *Nicotiana* and maize (Edqvist and Bergman 2002, Newton *et al.* 1995). While the identity of these factors is still unknown, mtTFBs belonging to the family of S-adenosyl-L-methionine (SAM)-dependent rRNA adenine dimethyltransferases were found to be essential for transcription in yeast and mammalian mitochondria (Asin-Cayuela and Gustafsson 2007, Jang and Jaehning 1991). The nuclear genome of *Arabidopsis thaliana* encodes three dimethylase/mtTFB-like proteins, one of which, Dim1B, has been shown to be imported into mitochondria (Richter *et al.* 2009). However, transcription initiation by mitochondrial RNA polymerases was not stimulated by Dim1B *in vitro* (Kühn 2006, Richter *et al.* 2009). In the light of proposed elements within the transcribed regions controlling transcription of a specific set of mitochondrial genes and a possible involvement of *trans* factors, it appears worthwhile to revaluate the design of the *in vitro* experiments. Possibly the most persuasive argument for the proposed model takes into account that due to frequent rearrangements of mitochondrial sequences, the mtDNAs of dicotyledonous plant species show occasional substitutions of promoters between different mitochondrial genes (Forner *et al.* 2008, Handa 2003, Holec *et al.* 2008b, Kubo *et al.* 2000, Unseld *et al.* 1997). Therefore, it might seem plausible that the transcription of mitochondrial genes could partly be controlled through elements within the transcribed regions rather than upstream of the TSS (Kühn *et al.* 2009).

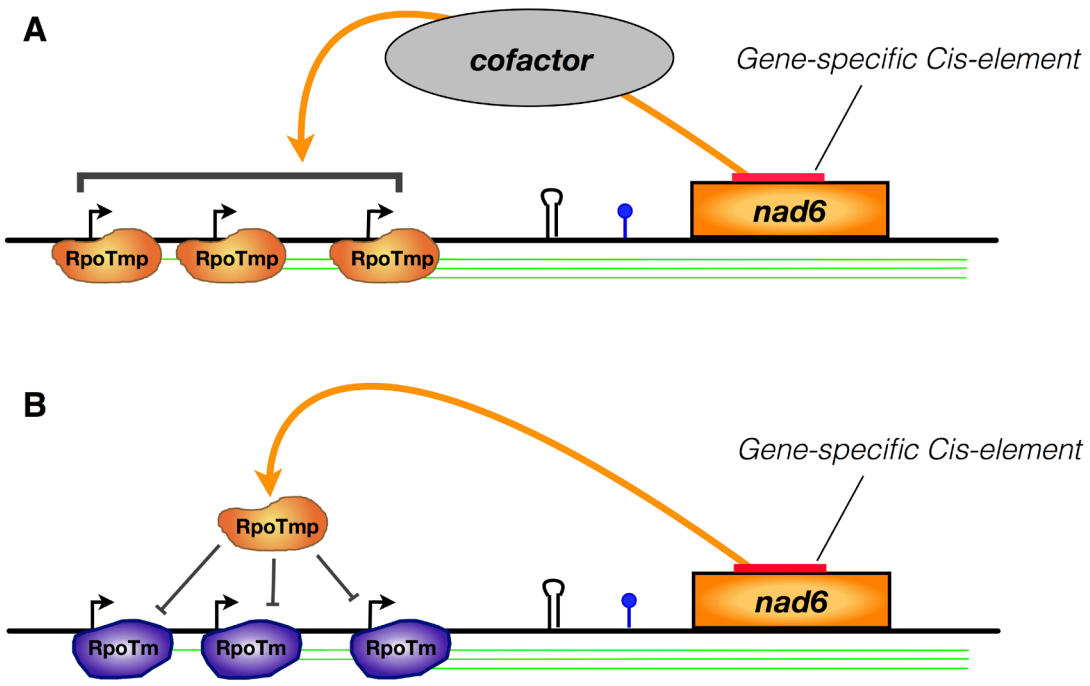


Figure 37: Two alternative models of AtRpoTmp dependent mitochondrial transcription.

4.5.4 Roles of RpoTmp- and RpoTm-dependent transcriptional mechanisms

The reported developmentally limited role of RpoTmp in plastids (see above) is in stark contrast with the importance of this enzyme for mitochondrial gene expression from seedlings to adult plants. RpoTm seems to function as the basic RNA polymerase in mitochondria, unconditionally required for the transcription of most or all mitochondrial genes. RpoTmp likely ameliorates this basic level by strongly enhancing transcription of a defined set of mitochondrial genes. Therefore, RpoTmp activity, apparently important for the expression of sufficient levels of respiratory complexes I and IV, might be a constituent that allows mitochondria to control the expression of a subset of mitochondrial genes independently and possibly in response to the developmental or metabolic requirements of the organelle and the cell. What are conditions where such a mechanism might operate? Under conditions of high ATP due to photosynthetic ATP production, the proton pumping respiratory complex I can be bypassed by substitute pathways, which allows for NADH reoxidation and in turn operation of the TCA (Fernie *et al.* 2004, Noguchi and Yoshida 2008, Rasmusson *et al.* 2008). Another alternative pathway is established through AOX a second terminal oxidase in mitochondria, which transfers electrons directly from reduced ubiquinol to oxygen forming water, thereby bypassing the proton pumping respiratory complex IV. These two examples likely highlight only a minor fraction of the highly complex adaptations of the plant respiratory chain to specific needs of photoautotrophic organisms. It is tempting to speculate that conditions might exist where cells regulate the abundances of respiratory chain complexes I and IV *via* mitochondrial transcription to meet their specific metabolic needs.

4.6 Antisense transcription and MTERF in mammals - implications for plant mitochondrial transcription

In mice MTERF3 has been recently shown to be a negative regulator of mtDNA transcription initiation (Park *et al.* 2007). Interestingly, loss of MTERF3 was shown to be embryonically lethal, and tissue-specific knockout of Mterf3 in the heart of mice resulted in increased transcription initiation on both strands of mtDNA, but to decreased steady-state levels of promoter-distal transcripts, and as a result to mitochondrial dysfunction. Therefore, the authors suggest that MTERF3 function provides a mechanism to avoid collision between transcription complexes generated by the opposing promoters (LSP, HSP) of mammalian mitochondrial genomes. Such a 'collision-model' is gaining weight since studies of yeast nuclear transcription have demonstrated that opposed closely spaced promoters have decreased steady-state levels of promoter-distal transcripts, even though transcription initiation was shown not to be affected (Prescott and Proudfoot 2002). Moreover the yeast studies showed decreased transcription initiation at the forward promoter to result in increased steady-state levels of promoter-distal transcripts initiated from the reverse promoter (Prescott and Proudfoot 2002). Direct evidence for such processes to occur in plant organelles is lacking to date, but asymmetric signal across specific transcripts in plastidial reverse run-on transcription experiments, with lower activity in promoter distal regions has been observed (Dr. Zubo, HU-Berlin pers. communication) and could be explained by collision events occurring, due to transcription from the opposite strand.

The established model of transcription in plant mitochondrial and plastid genomes, where multiple RNA polymerase activities are actively transcribing from multiple and opposing promoters on both strands, calls for the consideration of such effects, specifically when initiation of transcription is altered by inactivation of components of the transcriptional apparatus. Such a model might explain the puzzling observation of additional transcriptional start points in mutants that lack one RNA polymerase as observed in *PprpoTmp2* and *AtrpoTmp* plants. While in yeast and in the mouse studies transcription was facilitated by the same RNAP the situation in *Arabidopsis* and *Physcomitrella* organelles is significantly different. Multiple polymerases with different properties have to be considered to influence each other and in turn transcriptional performance. For example, in *Physcomitrella* wild-type mitochondria elements within or downstream from the transcribed regions, functioning as antisense promoters utilised by PpRpoTmp2, could repress transcription from forward promoters and therefore explain the observed elevation of steady state transcript levels and the identification of additional promoters in *PprpoTmp2* plants. However, in *Arabidopsis* such a model is lacking support, since elevation of steady state transcript levels is accompanied by the elevation of transcription initiation. Nevertheless future studies on the function of RpoT proteins in both organelles should consider possible interactions of multiple RNAPs, such as interaction in transcriptional initiation and elongation (Fig. 37B). Of note, mammalian DNA polymerase core proteins for instance are suggested to be exchanged in the process of the elongation phase (pers. communication Dr. Euro, University of Helsinki). To gain a comprehensive picture of the organellar RNA metabolism in wild-type and RpoT mutant plants deep sequencing of the transcriptome and the initiome should be considered. The high resolution and strand specificity of such an approach could provide additional inside on the function of RpoT enzymes and the gene expression machinery in general.

Furthermore new studies of the AtRpoTmp *in vitro* activity and more importantly complementation studies with putative transcription factor proteins are needed and should be modified in such a way that DNA templates used in the assay contain not only the sequences

surrounding the identified TSS (Richter *et al.* 2009), but the downstream transcribed regions as well. Additionally it has been shown that the secondary structure of a DNA single strand confers activation of N4 bacteriophage promoters dependent on the unwinding and stabilisation of the DNA by gyrase and SSB (Davydova *et al.* 2007). N4 is a member of the T7-like RNAP family and both *E. coli* proteins have functional equivalents in plants reported to be targeted to mitochondria and chloroplasts (Wall *et al.* 2004, Zaegel *et al.* 2006). A putative role of both plant proteins in transcriptional initiation by possibly exposing a hidden conservation not recognisable in the primary promoter structure could also be tested *in vitro*.

4.7 Evolution of monomeric *Superfamily A* like RNA polymerases

The data of the analysis presented herein are supportive of the view that the ssRNAP gene originated early in evolutionary history before the endosymbiotic origin of present day mitochondria, possibly as suggested only recently from the newly found TV-Pols (***T***ransposon***V***irus–Pols). More precisely, the three structural units found in ssRNAPs, finger, palm, thumb have been proposed to originate from different polypeptides, the fusion of which resulted in the typical *Superfamily A* polymerases structure. Specifically, the newly found TV-Pols with predicted primase polymerase activity are suggested to possibly represent an ancestral viral *Superfamily A* polymerase that might have contributed to the evolution of ssRNAPs (Iyer *et al.* 2008).

For the origin of phage-type RNA polymerases in eukaryotes Shutt and Gray proposed, that a T-Odd prophage (T7-group of *Podoviridae*) present in the ancestral α–proteobacterium at the origin of mitochondria supplied the RNA polymerase, which replaced the original bacterial enzymes of the mitochondrial endosymbiont. Even though the scenario gained weight due to the massive sequencing of bacterial and phage genomes bearing witness of nonlytic T7-like podoviral phages in protobacterial genomes and suggestive of a broader host range of these phages, phylogenetic evidence neither supports a specific common ancestry of mitochondrial plasmid- encoded with nucleus-encoded ssRNAP nor of the latter with phage RNAPs. Moreover the limited phylogenetic distribution of plasmid encoded ssRNAP genes found only in certain angiosperms and fungi, together with strong evidence for horizontal transfer events, disfavour the view of mitochondrial linear plasmids as a transitional form in the acquisition of an ssRNAP gene by the nuclear genome. Nevertheless, an early origin scenario prior to or at the separation of the three domains of life gained support due to the observation of a host range beyond γ-proteobacteria of T7-like podoviral phages allowing for the assumption of ssRNAP gene sequestration in an independent infectious entity accounting for the distribution of ssRNAPs observed in phages and eukaryotes. Data provided herein are in support of the hypothesis, that present day mitochondria facilitate 'T3/T7- like' proteins because a cryptic prophage was integrated into the genome of the ancestral α-proteobacterial endosymbiont, thereby challenging two alternative scenarios for the evolutionary source of the nucleus-encoded ssRNAPs. Nucleus-encoded ssRNAPs were likely neither present in the nuclear genome of the pre-eukaryotic cell that served as host for the mitochondrial endosymbiosis nor contributed by a plasmid-like agent.

Phylogenetic analysis (Figure 31) yielded only two well-defined clusters, due to the nesting of the plasmid encoded proteins inside the phage cluster. Monophyly of the nucleus-encoded

proteins suggests a single transfer event as the origin of the mitochondrial targeted ssRNAPs in eukaryotes. The mitochondrial plasmid-encoded sequences, though nested in the phage cluster appear as a monophyletic lineage. While the nested position makes these genes a unlikely 'source' of the nuclear encoded mitochondrial ssRNAPs, monophyly is in stark contrast to the observed phylogenetic scattering of the members in the cluster. Long branch attraction could certainly be a reason for the observed pattern. However, a biological explanation might also be possible. Together with the scattered phylogenetic distribution of the plasmid encoded ssRNAPs in eukaryotes, the observed pattern could reflect the existence of phylogenetically closely related cryptic vectors, which by themselves account for the monophyletic pattern. Accordingly, the plasmid encoded ssRNAPs would be the result of multiple independent HGT events. Such a scenario could also explain why a continuous lineage for the plasmid ssRNAPs is neither observed in plants nor in fungi. Both plant and fungal linear plasmids usually contain terminal inverted repeats and encode DNA and/or RNA polymerases possibly to drive expression and replication in an autonomous manner. This invertron structure also found in some DNA viruses like *phi29* and adenoviruses suggests a viral origin of plant and fungal linear plasmids. Furthermore, structural variations between plant linear plasmids and their unbalanced distribution strongly indicate diverse origins rather than common ancestry inside of the plant kingdom. To date the most likely scenario is that linear plasmids present in only nine plant species, were acquired multiple times via horizontal gene transfer, possibly involving fungal phyto-pathogens (Handa 2008, McDermott *et al.* 2008). Thus fungal targets of a virus-like entity could also be the source of the mitochondrial plasmids encoding ssRNAP ORFs found in a limited number of plant species.

The distribution of prophage single enzyme RNAP ORFs inside the protobacterial phylum, shows an interesting pattern. Five coding frames were found in *α-proteobacteria*, four in β-, three in *γ-proteobacteria* and one in δ-*proteobacteria*. Considering the bias of available fully sequenced proteobacterial genomes, three times as many γ-proteobacterial genomes are sequenced to date, the number of α-proteobacterial genomes containing ORFs for single enzyme RNAPs is most likely significantly higher when compared with *γ-proteobacteria*. This would in turn change the view of podoviral phages as mainly infecting *γ-proteobacteria*. While it was already reported, that the host-specificity of this group of phages is likely broader than previously thought (Filee and Forterre 2005, Shutt and Gray 2006a), with the extreme of autographivirinal phages infecting cyanobacteria, these new observations could be seen as circumstantial evidence for a high number of thus far undetected phages with monomeric DNA dependent RNA polymerases infecting α-proteobacteria, the subdivision from which mitochondria derive.

A relatively early origin of monomeric RNAPs is consistent with the shown, wide phylogenetic distribution of *ssRNAP* sequences and the extent of sequence divergence among the three classes of *ssRNAP* genes, i.e. phage, mitochondrial plasmid and nuclear eukaryotic ORFs (Cermakian *et al.* 1997). An early origin scenario as early as prior to the separation of the three domains of life, or even paralleling early eubacterial evolution due to the sequestration of this gene in an independent infectious entity, such as a bacteriophage, seems most likely, and in line with the wide host range, of ssRNAP carrying bacteriophages. Therefore, evolution of podoviral and/or plasmid-encoded ssRNAPs from an original nucleus-encoded ssRNAP is rather unlikely.

Considering a subsequent gene transfer from mitochondria to the nucleus, a prophage that might have been part of the α-proteobacterial endosymbionts genome or an active podoviral phage, present in the α-proteobacterial ancestor of mitochondria at the time of the endosymbiotic uptake, is therefore the most likely source for nuclear encoded phage-type RNA polymerases. A predicament for all hypotheses, concerning the origin of phage-type RNA polymerases, is the observation that the mitochondrial genome of the jacobide *Reclinomonas americana* (see 1.2.1) still contains all genes for the bacterial-type RNA polymerase lost in all other eukaryotic lineages investigated thus far. It also represents a major problem for the rooting of the eukaryotic tree, considering that jacobids are not basal to all other eukaryotes, but seem to belong to the monophyletic group of jacobids, *Euglenozoa* plus *Heterolobosea* (JEH group) with *Euglenozoa*, *Heterolobosea* known to encode T7-like polymerases in their nuclear genomes (Rodriguez-Ezpeleta *et al.* 2007, Simpson *et al.* 2006). Notably it is still not clear, whether *Reclinomonas* encodes a phage-type polymerase in the nuclear genome or uses the mitochondrial encoded bacterial-type polymerase exclusively (Shutt and Gray 2006a). Therefore, it seems possible to suggest that phage-type RNA polymerases and the original bacterial-type enzymes might have coexisted in various lineages over a prolonged period, the bacterial-type polymerase being lost differentially and as a result only found in *Reclinomonas americana* (Rodriguez-Ezpeleta *et al.* 2007). This scenario is especially tempting, since – besides avoiding homoplasy – a very similar pattern is observable in chloroplasts of land plants. Both bacterial- and phage-type activities are utilised in transcription and even differential loss, as reported for parasitic plants, is observed (Krause 2008).

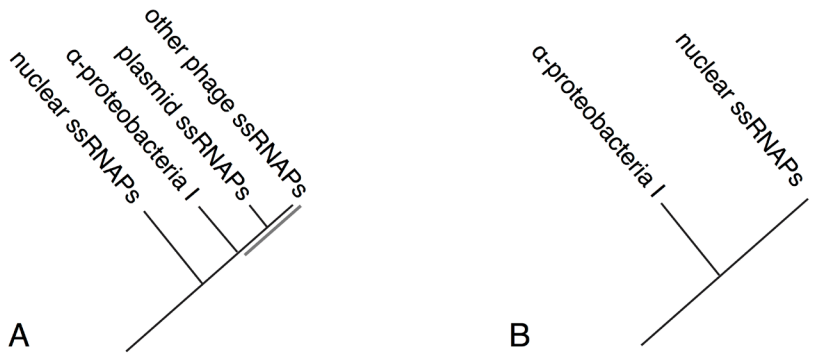


Figure 38: Relation of nuclear encoded mitochondrial RNA polymerases to other monomeric RNAP proteins. Rooting on any of the branches indicated by the grey bar in A result in the terminal dichotomy shown in B

Assuming a bacteriophage origin of ssRNAPs all recent podoviral ssRNAPs would be orthologous to the nuclear encoded mitochondrial enzymes, if the phylogenetic tree is rooted as shown in Figure 38A. Even though such a scenario cannot be dismissed, the origin of ssRNAPs in a bacterial virus-like entity might favour the rooting inside the phage cluster. This in turn would result in a terminal dichotomy of nuclear encoded eukaryotic and phage polymerases from two alphaproteobacterial prophage insertions (Fig. 38B). In line with the α-proteobacterial origin of mitochondria it would directly substantiate the scenario proposed by Shutt and Gray, where in a single event a phage, infecting α-proteobacteria, possibly as a prophage present in the protomitochondrial genome gave rise to nuclear encoded phage-type

polymerases of eukaryotes, with nuclear localisation attributable to a secondary transfer of the gene to the nucleus.

Thus far multiple ssRNAPs were only observed in land plants. Therefore, one fascinating additional observation was the identification of eight fungal species each encoding two putative phage-type RNA Polymerases in their nuclear genomes.

The origin of the second ssRNAP is likely due to a single horizontal gene transfer (HGT) event from a deep branching eukaryote in a direct ancestor of *mitosporic Saccharomycetales* and *Debaryomycetaceae.*

Interestingly, this subgroup of Saccharomycetales is characterised by other unusual features. Reportedly, species of this cluster translate CTG codons as serine rather than leucine, a reassignment that occurred more than 170 million years ago (Massey *et al.* 2003, Miranda *et al.* 2006). As shown for *Debaryomyces hansenii* massive integrations of NUMTs (**nu**clear copies of **mi**t**o**chondrial DNA) and NUPAVs (**nu**clear sequences of **pla**smid and **v**iral origin) are found in its nuclear genome, speculated to be the consequence of the use of promiscuous DNA fragments in double strand break repair processes at fragile chromosomal sites (Frank and Wolfe 2009). Possibly, these unusual features might have fostered the acquisition of a second ssRNAP gene. The identification of two ssRNAP coding frames in some fungi suggests, that multiple phage-type RNA polymerases might not be a unique feature of land plant species. Whereas in the relatively small streamlined mitochondrial genome of metazoans transcription is initiated from one or two uni- or bidirectional promoters in the single D-loop, the region that also contains most regulatory elements for replication and transcription (Tracy and Stern 1995), in fungal and plant mitochondrial genomes multiple promoters are facilitated (Kühn *et al.* 2005, Notsu *et al.* 2002, Sugiyama *et al.* 2005, Unseld *et al.* 1997). Therefore, fungal mitochondrial genomes in this sense are more plant like than their metazoan pendants. Considering the specific functions of the two mitochondrial RNAPs in *Arabidopsis* described herein, it is of great interest to determine the sub-cellular localisation of both proteins and to ascertain if a similar division of labour evolved in case of the two putative phage-type RNA polymerases from *mitosporic Saccharomycetales* and *Debaryomycetaceae*.

4.8 Evolution of RpoT proteins in Viridiplantae

All plant proteins used for the phylogenetic reconstruction of plant phage-type RNA polymerases contain the characteristic domains and all residues shown to be essential for the enzymatic activity of ssRNAPs. The intron-exon structure of land plant *RpoT* genes has been shown to be highly conserved from mosses to flowering plants (Hedtke *et al.* 1997, Richter *et al.* 2002, Yin *et al.* 2009, Yin *et al.* 2010) indicating that all land plant RpoT sequences decent from a common ancestral RpoT gene. Phylogenetic trees were reconstructed to revaluate the molecular phylogeny of the RpoT polymerases in the plant kingdom and to reassess the phylogenetic positioning of the polymerases characterised in the present study.

In the phylogenetic analysis angiosperm RpoT polymerases are found to constitute two clearly discernible groups with one consisting of plastid localised polymerases, and the other

of mitochondrial-localised and dual-targeted enzymes. The data presented suggest that all RpoTp proteins descent from a common duplication event that took place in a common ancestor of all flowering plants. Thus far it is unknown whether ferns or gymnosperms contain nuclear genes encoding plastid-localised phage-type RNAPs as well. Since the duplication event giving rise to the second NEP activity in eudicots is clearly more recent, identification of a purely plastid-localised phage-type RNAP in the basal angiosperm *Nuphar advena*, orthologous to all other purely plastid-targeted enzymes (RpoTp) of flowering plants, suggests that the acquisition of a nuclear gene-encoded transcriptional activity for plastids, not present in lycopods, took place after the split of lycopods from all other tracheophytes, with or before the rise of flowering plants. Moreover, the lack of a dual-targeted RpoTmp both in *Nuphar* and in monocots indicates that the RpoTmp enzyme detected in eudicots is an evolutionary novelty due to an *RpoTm* gene duplication that likely occurred only after the separation of monocots and eudicots. The putative plastid targeting sequences as present in two of the three *Physcomitrella* RpoT proteins are therefore clearly species- or lineage-specific convergent inventions. Importantly, multiple mitochondrial RNA polymerases as found in *Physcomitrella* and eudicots are indentified in *Nuphar* as well. The fixation of duplicated *RpoT* genes led to a convergent multiplicity of mitochondrial RNAPs in *Nuphar*, *Physcomitrella* and eudicots, previously not found in any other eukaryotic lineage.

Whereas *Pinus*, *Selaginella* and *Physcomitrella* polymerases do not belong to the branches of well separated plastid and mitochondrial (and dual targeted) polymerases, the RpoT polymerases from the basal angiosperm *Nuphar advena* cluster with the branches of plastid and mitochondrial/dual targeted sequences: NaRpoTm1 and NaRpoTm2 within the mitochondrial, and NaRpoTp within the plastid branch. The positioning of the RpoT protein from *Pinus taeda* with putative mitochondrial targeting properties (Beick et al., unpublished data) on the tree is in agreement with the land plant phylogeny (Finet *et al.* 2010), but the identification of thus far only one *RpoT* gene in this gymnosperm must be taken cautiously since the huge genome size (21Gb) is obstructing the screening for additional *RpoT* genes.

The presence of only one *RpoT* gene in *Selaginella* (Yin *et al.* 2009) contrasts the situation in the leafy moss *P. patens* and flowering plants. Thus, the lycophyte *Selaginella* encodes only the indispensable mitochondrial phage-type RNA polymerase. Another pecularity of SmRpoTm is its position on the tree. In the phylogenetic tree of plants, Lycophytes establish a sister group to the Euphyllophyta, which comprise seed plants and ferns, while leafy mosses like *Physcomitrella*, as constituents of a deeper branch are a sister group of possibly hornworth, but well established lycophytes, and all other tracheophytes (Finet *et al.* 2010). Therefore, it was surprising that the *Selaginella* protein incongruent with the well established land plant phylogeny appears as a paralog of all other land plant RpoT proteins including mosses. Since bootstrap support is sufficiently high, one has to assume an ancestral state of at least two RpoT proteins in the common ancestor of bryophytes and later branching land plants. Accordingly, the orthologue of the *Selaginella* RpoT protein paralogous to all other land plant RpoTs was lost twice independently in the evolution of land plants: firstly in the *Physcomitrella* lineage and secondly after the split of the *Selaginella* lineage from all higher branching tracheophytes. The orthologue of all other land plant RpoT proteins was lost in the *Selaginella* lineage. Therefore, possession of only one RpoT polymerase as observed in *Selaginella* rather seems to be a derived state. All *Physcomitrella* RpoTs on the other hand appear to be the result of two *Physcomitrella* lineage specific duplication events, with the

PpRpoTm/RpoTmp1 duplication as the more recent one. This might indicate conservation the plastid targeting properties for both dual targeted proteins, or alternatively that the plastid targeting properties have been acquired twice independently. In either of those two scenarios the functional importance seems to be strongly indicated.

Multiple RpoT activities per organelle are a common feature of eudicotyledonous plants, *Nuphar advena* (only in mitochondria) and *Physcomitrella patens*, but are apparently absent from monocots. In plastids of most land plants the complexity of the transcriptional apparatus is even higher due to the plastid encoded multi-subunit RNAP. Whether this complexity is due to an evolutionary trend adaptive or neutral (see above) remains open. The identification of different but overlapping functions of RpoT proteins in plastids as well as in mitochondria of *Arabidopsis* is only the first step to answer such a question, since a regulatory role *sensu stricto* for these enzymes remains to be shown. In *Physcomitrella patens* the data provided herein are suggestive of divergent but overlapping functions of RpoT enzymes as well but direct proof for a specific function of PpRpoTmp2 is still lacking. The specific functions of RpoTmp2 and RpoTm deserve further attention and studies applying whole transcriptome/ initiome analysis to both organelles could greatly enhance our knowledge of the transcriptional machineries in mitochondria and plastids. If the RNAi method could be established in *Nuphar advena*, this system might offer a way to show if a similar division of labour evolved in case of the two mitochondrial RNA polymerases in this plant, offering another perspective on the complex and possibly adaptive nature of facilitating multiple RpoTs in mitochondrial transcription.

5 References

Allison, L.A. (2000) The role of sigma factors in plastid transcription. *Biochimie*, **82**, 537-548.

Allison, L.A., Simon, L.D. and Maliga, P. (1996) Deletion of rpoB reveals a second distinct transcription system in plastids of higher plants. *EMBO J*, **15**, 2802-2809.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457-465.

Aravind, L., Anantharaman, V. and Iyer, L.M. (2003) Evolutionary connections between bacterial and eukaryotic signaling systems: a genomic perspective. *Curr Opin Microbiol*, **6**, 490-497.

Asin-Cayuela, J. and Gustafsson, C.M. (2007) Mitochondrial transcription and its regulation in mammalian cells. *Trends Biochem Sci*, **32**, 111-117.

Azevedo, J., Courtois, F., Hakimi, M.A., Demarsy, E., Lagrange, T., Alcaraz, J.P., Jaiswal, P., Marechal-Drouard, L. and Lerbs-Mache, S. (2008) Intraplastidial trafficking of a phage-type RNA polymerase is mediated by a thylakoid RING-H2 protein. *Proc Natl Acad Sci U S A*, **105**, 9123-9128.

Baba, K., Schmidt, J., Espinosa-Ruiz, A., Villarejo, A., Shiina, T., Gardestrom, P., Sane, A.P. and Bhalerao, R.P. (2004) Organellar gene transcription and early seedling development are affected in the rpoT;2 mutant of Arabidopsis. *Plant J*, **38**, 38-48.

Backert, S., Dorfel, P. and Börner, T. (1995) Investigation of plant organellar DNAs by pulsed-field gel electrophoresis. *Curr Genet*, **28**, 390-399.

Baeza, L., Bertrand, A., Mache, R. and Lerbs-Mache, S. (1991) Characterization of a protein binding sequence in the promoter region of the 16S rRNA gene of the spinach chloroplast genome. *Nucleic Acids Res*, **19**, 3577-3581.

Barkan (1988) Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO*, **8**, 2637-2644.

Barnes, W.M. and Tuley, E. (1983) DNA sequence changes of mutations in the histidine operon control region that decrease attenuation. *J Mol Biol*, **165**, 443-459.

Beardslee, T.A., Roy-Chowdhury, S., Jaiswal, P., Buhot, L., Lerbs-Mache, S., Stern, D.B. and Allison, L.A. (2002) A nucleus-encoded maize protein with sigma factor activity accumulates in mitochondria and chloroplasts. *Plant J*, **31**, 199-209.

Bedbrook, J.R. and Bogorad, L. (1976) Endonuclease recognition sites mapped on Zea mays chloroplast DNA. *Proc Natl Acad Sci U S A*, **73**, 4309-4313.

Bendich, A.J. (2004) Circular chloroplast chromosomes: the grand illusion. *Plant Cell*, **16**, 1661-1666.

Bensing, B.A., Meyer, B.J. and Dunny, G.M. (1996) Sensitive detection of bacterial transcription initiation sites and differentiation from RNA processing sites in the pheromone-induced plasmid transfer system of Enterococcus faecalis. *Proc Natl Acad Sci U S A*, **93**, 7794-7799.

Berry, S. (2003) Endosymbiosis and the design of eukaryotic electron transport. *Biochim Biophys Acta*, **1606**, 57-72.

Binder, S. and Brennicke, A. (2003) Gene expression in plant mitochondria: transcriptional and post-transcriptional control. *Philos Trans R Soc Lond B Biol Sci*, **358**, 181-188; discussion 188-189.

Binder, S., Marchfelder, A. and Brennicke, A. (1996) Regulation of gene expression in plant mitochondria. *Plant Mol Biol*, **32**, 303-314.

Biswas, T.K. (1999) Nucleotide sequences surrounding the nonanucleotide promoter motif influence the activity of yeast mitochondrial promoter. *Biochemistry*, **38**, 9693-9703.

Biswas, T.K., Edwards, J.C., Rabinowitz, M. and Getz, G.S. (1985) Characterization of a yeast mitochondrial promoter by deletion mutagenesis. *Proc Natl Acad Sci U S A*, **82**, 1954-1958.

Biswas, T.K. and Getz, G.S. (1986a) A critical base in the yeast mitochondrial nonanucleotide promoter. Abolition of promoter activity by mutation at the -2 position. *J Biol Chem*, **261**, 3927-3930.

Biswas, T.K. and Getz, G.S. (1986b) Nucleotides flanking the promoter sequence influence the transcription of the yeast mitochondrial gene coding for ATPase subunit 9. *Proc Natl Acad Sci U S A*, **83**, 270-274.

Biswas, T.K., Ticho, B. and Getz, G.S. (1987) In vitro characterization of the yeast mitochondrial promoter using single-base substitution mutants. *J Biol Chem*, **262**, 13690-13696.

Bligny, M., Courtois, F., Thaminy, S., Chang, C.C., Lagrange, T., Baruah-Wolff, J., Stern, D. and Lerbs-Mache, S. (2000) Regulation of plastid rDNA transcription by interaction of CDF2 with two different RNA polymerases. *EMBO J*, **19**, 1851-1860.

Bogenhagen, D.F. (1996) Interaction of mtTFB and mtRNA polymerase at core promoters for transcription of Xenopus laevis mtDNA. *J Biol Chem*, **271**, 12036-12041.

Bogenhagen, D.F. and Insdorf, N.F. (1988) Purification of Xenopus laevis mitochondrial RNA polymerase and identification of a dissociable factor required for specific transcription. *Mol Cell Biol*, **8**, 2910-2916.

Bohne A.V., Weihe A. and Börner, T. (2009) Transfer RNAs inhibit Arabidopsis phage-type RNA polymerases. *Endocytobiosis Cell Res. (2009) 19, 63-69*, 1-7.

Bonen, L. and Vogel, J. (2001) The ins and outs of group II introns. *Trends Genet*, **17**, 322-331.

Boyer, S.K. and Mullet, J.E. (1988) Sequence and transcript map of barley chloroplast psbA gene. *Nucleic Acids Res*, **16**, 8184.

Bradley, D. and Gatenby, A.A. (1985) Mutational analysis of the maize chloroplast ATPase-beta subunit gene promoter: the isolation of promoter mutants in E. coli and their characterization in a chloroplast in vitro transcription system. *EMBO J*, **4**, 3641-3648.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. and Schuster, W. (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett*, **325**, 140-145.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C. (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol*, **18**, 225-239.

Burger, G., Gray, M.W. and Lang, B.F. (2003) Mitochondrial genomes: anything goes. *Trends Genet*, **19**, 709-716.

Burzio, V.A., Villota, C., Villegas, J., Landerer, E., Boccardo, E., Villa, L.L., Martinez, R., Lopez, C., Gaete, F., Toro, V., Rodriguez, X. and Burzio, L.O. (2009)

Expression of a family of noncoding mitochondrial RNAs distinguishes normal from cancer cells. *Proc Natl Acad Sci U S A*, **106**, 9430-9434.

Cahoon, A.B., Harris, F.M. and Stern, D.B. (2004) Analysis of developing maize plastids reveals two mRNA stability classes correlating with RNA polymerase type. *EMBO Rep*, **5**, 801-806.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W. and Cedergren, R. (1997) On the evolution of the single-subunit RNA polymerases. *J Mol Evol*, **45**, 671-681.

Chang, C.C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S. and Stern, D.B. (1999) Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. *Plant Cell*, **11**, 911-926.

Chateigner-Boutin, A.L. and Hanson, M.R. (2002) Cross-competition in transgenic chloroplasts expressing single editing sites reveals shared cis elements. *Mol Cell Biol*, **22**, 8448-8456.

Cheetham, G.M. and Steitz, T.A. (2000) Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr Opin Struct Biol*, **10**, 117-123.

Cheng, Y.S., Lin, C.H. and Chen, L.J. (1997) Transcription and processing of the gene for spinach chloroplast threonine tRNA in a homologous in vitro system. *Biochem Biophys Res Commun*, **233**, 380-385.

Cho, Y. and Palmer, J.D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol Biol Evol*, **16**, 1155-1165.

Clifton, R., Lister, R., Parker, K.L., Sappl, P.G., Elhafez, D., Millar, A.H., Day, D.A. and Whelan, J. (2005) Stress-induced co-expression of alternative respiratory chain components in Arabidopsis thaliana. *Plant Mol Biol*, **58**, 193-212.

Cotney and Shadel (2006) Evidence for an Early Gene Duplication Event in the Evolution of the Mitochondrial Transcription Factor B Family and Maintenance of rRNA Methyltransferase Activity in Human mtTFB1 and mtTFB2. *J Mol Evol*, **63**, 707–717.

Courtois, F., Merendino, L., Demarsy, E., Mache, R. and Lerbs-Mache, S. (2007) Phage-type RNA polymerase RPOTmp transcribes the rrn operon from the PC promoter at early developmental stages in Arabidopsis. *Plant Physiol*, **145**, 712-721.

Davydova, E.K., Santangelo, T.J. and Rothman-Denes, L.B. (2007) Bacteriophage N4 virion RNA polymerase interaction with its promoter DNA hairpin. *Proc Natl Acad Sci U S A*, **104**, 7033-7038.

de Longevialle, A.F., Hendrickson, L., Taylor, N.L., Delannoy, E., Lurin, C., Badger, M., Millar, A.H. and Small, I. (2008) The pentatricopeptide repeat gene OTP51 with two LAGLIDADG motifs is required for the cis-splicing of plastid ycf3 intron 2 in Arabidopsis thaliana. *Plant J*, **56**, 157-168.

de Longevialle, A.F., Meyer, E.H., Andres, C., Taylor, N.L., Lurin, C., Millar, A.H. and Small, I.D. (2007) The pentatricopeptide repeat gene OTP43 is required for trans-splicing of the mitochondrial nad1 Intron 1 in Arabidopsis thaliana. *Plant Cell*, **19**, 3256-3265.

Deng, X.W., Wing, R.A. and Gruissem, W. (1989) The chloroplast genome exists in multimeric forms. *Proc Natl Acad Sci U S A*, **86**, 4156-4160.

Dombrowski, S., Brennicke, A. and Binder, S. (1997) 3'-Inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *EMBO J*, **16**, 5069-5076.

Dombrowski, S., Hoffmann, M., Guha, C. and Binder, S. (1999) Continuous primary sequence requirements in the 18-nucleotide promoter of dicot plant mitochondria. *J Biol Chem*, **274**, 10094-10099.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T. and Wilson, A. (2010) Geneious v5.1, available from http://www.geneious.com/.

Drummond, A., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T. and Wilson, A. (2008) Geneious v4.0, available from http://www.geneious.com/.

Duchene, A.M. and Marechal-Drouard, L. (2001) The chloroplast-derived trnW and trnM-e genes are not expressed in Arabidopsis mitochondria. *Biochem Biophys Res Commun*, **285**, 1213-1216.

Dutilleul, C., Driscoll, S., Cornic, G., De Paepe, R., Foyer, C.H. and Noctor, G. (2003) Functional mitochondrial complex I is required by tobacco leaves for optimal photosynthetic performance in photorespiratory conditions and during transients. *Plant Physiol*, **131**, 264-275.

Edqvist, J. and Bergman, P. (2002) Nuclear identity specifies transcriptional initiation in plant mitochondria. *Plant Mol Biol*, **49**, 59-68.

Eisermann, A., Tiller, K. and Link, G. (1990) In vitro transcription and DNA binding characteristics of chloroplast and etioplast extracts from mustard (*Sinapis alba*) indicate differential usage of the psbA promoter. *EMBO J*, **9**, 3981-3987.

Emanuel, C., von Groll, U., Muller, M., Börner, T. and Weihe, A. (2006) Development- and tissue-specific expression of the RpoT gene family of Arabidopsis encoding mitochondrial and plastid RNA polymerases. *Planta*, **223**, 998-1009.

Emanuel, C., Weihe, A., Graner, A., Hess, W.R. and Börner, T. (2004) Chloroplast development affects expression of phage-type RNA polymerases in barley leaves. *Plant J*, **38**, 460-472.

Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, **300**, 1005-1016.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N. and Gustafsson, C. (2002) Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat Genet*, **31**, 289-294.

Falkenberg, M., Larsson, N.G. and Gustafsson, C.M. (2007) DNA replication and transcription in mammalian mitochondria. *Annu Rev Biochem*, **76**, 679-699.

Fernie, A.R., Carrari, F. and Sweetlove, L.J. (2004) Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol*, **7**, 254-261.

Fey, J. and Marechal-Drouard, L. (1999) Compilation and analysis of plant mitochondrial promoter sequences: An illustration of a divergent evolution between monocot and dicot mitochondria. *Biochem Biophys Res Commun*, **256**, 409-414.

Filee, J. and Forterre, P. (2005) Viral proteins functioning in organelles: a cryptic origin? *Trends Microbiol*, **13** (11), 510-512.

Finet, C., Timme, R.E., Delwiche, C.F. and Marletaz, F. (2010) Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Curr Biol*, **20**, 2217-2222.

Finnegan, P.M. and Brown, G.G. (1990) Transcriptional and Post-Transcriptional Regulation of RNA Levels in Maize Mitochondria. *Plant Cell*, **2**, 71-83.

Fischer, N., Stampacchia, O., Redding, K. and Rochaix, J.D. (1996) Selectable marker recycling in the chloroplast. *Mol Gen Genet*, **251**, 373-380.

Fisher, R.P. and Clayton, D.A. (1988) Purification and characterization of human mitochondrial transcription factor 1. *Mol Cell Biol*, **8**, 3496-3509.

Forner, J., Holzle, A., Jonietz, C., Thuss, S., Schwarzlander, M., Weber, B., Meyer, R.C. and Binder, S. (2008) Mitochondrial mRNA polymorphisms in different Arabidopsis accessions. *Plant Physiol*, **148**, 1106-1116.

Forner, J., Weber, B., Thuss, S., Wildum, S. and Binder, S. (2007) Mapping of mitochondrial mRNA termini in Arabidopsis thaliana: t-elements contribute to 5' and 3' end formation. *Nucleic Acids Res*, **35**, 3676-3692.

Frank, A.C. and Wolfe, K.H. (2009) Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. *Eukaryot Cell*, **8**, 1521-1531.

Gagliardi, D. and Gualberto, J.M. (2004) Gene expression in higher plant mitochondria. In *Advances in Photosynthesis and Respiration* (Day, A.D., Millar, A.H. and Whelan, J. eds). Dorcrecht: Kluwer, 55-81.

Galli, G., Hofstetter, H. and Birnstiel, M.L. (1981) Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature*, **294**, 626-631.

Gardner, J.F. (1982) Initiation, pausing, and termination of transcription in the threonine operon regulatory region of Escherichia coli. *J Biol Chem*, **257**, 3896-3904.

Garmier, M., Carroll, A.J., Delannoy, E., Vallet, C., Day, D.A., Small, I.D. and Millar, A.H. (2008) Complex I dysfunction redirects cellular and mitochondrial metabolism in Arabidopsis. *Plant Physiol*, **148**, 1324-1341.

Gaspari, M., Falkenberg, M., Larsson, N.G. and Gustafsson, C.M. (2004) The mitochondrial RNA polymerase contributes critically to promoter specificity in mammalian cells. *EMBO J*, **23**, 4606-4614.

Gatenby, A.A., Castleton, J.A. and Saul, M.W. (1981) Expression in E. coli of maize and wheat chloroplast genes for large subunit of ribulose bisphosphate carboxylase. *Nature*, **291**, 117-121.

Geiduschek, E.P. and Kassavetis, G.A. (1995) Comparing transcriptional initiation by RNA polymerases I and III. *Curr Opin Cell Biol*, **7**, 344-351.

Georg, J., Honsel, A., Voss, B., Rennenberg, H. and Hess, W.R. (2010) A long antisense RNA in plant chloroplasts. *New Phytol*, **186**, 615-622.

Georg, J., Voss, B., Scholz, I., Mitschke, J., Wilde, A. and Hess, W.R. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol Syst Biol*, **5**, 305.

Giegé, P. and Brennicke, A. (2001) From gene to protein in higher plant mitochondria. *C R Acad Sci III*, **324**, 209-217.

Giegé, P., Hoffmann, M., Binder, S. and Brennicke, A. (2000) RNA degradation buffers asymmetries of transcription in Arabidopsis mitochondria. *EMBO Rep*, **1**, 164-170.

Giegé, P., Sweetlove, L.J., Cognat, V. and Leaver, C.J. (2005) Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in Arabidopsis. *Plant Cell*, **17**, 1497-1512.

Goffinet, B., Wickett, N.J., Shaw, A.J. and Cox, C.J. (2005) Phylogenetic significance of the rpoA loss in the chloroplast genome of mosses. *Taxon*, **54**, 353-360.

Gray, M.W., Burger, G. and Lang, B.F. (1999) Mitochondrial evolution. *Science*, **283**, 1476-1481.

Gray, M.W. and Lang, B.F. (1998) Transcription in chloroplasts and mitochondria: a tale of two polymerases. *Trends Microbiol*, **6**, 1-3.

Gray, M.W., Lang, B.F. and Burger, G. (2004) Mitochondria of protists. *Annu Rev Genet*, **38**, 477-524.

Greenberg, B.M., Narita, J.O., DeLuca-Flaherty, C., Gruissem, W., Rushlow, K.A. and Hallick, R.B. (1984) Evidence for two RNA polymerase activities in Euglena gracilis chloroplasts. *J Biol Chem*, **259**, 14880-14887.

Greenleaf, A.L., Kelly, J.L. and Lehman, I.R. (1986) Yeast RPO41 gene product is required for transcription and maintenance of the mitochondrial genome. *Proc Natl Acad Sci U S A*, **83**, 3391-3394.

Gruissem, W., Elsner-Menzel, C., Latshaw, S., Narita, J.O., Schaffer, M.A. and Zurawski, G. (1986) A subpopulation of spinach chloroplast tRNA genes does not require upstream promoter elements for transcription. *Nucleic Acids Res*, **14**, 7541-7556.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, **52**, 696-704.

Hajdukiewicz, P.T., Allison, L.A. and Maliga, P. (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *EMBO J*, **16**, 4041-4048.

Haley, J. and Bogorad, L. (1990) Alternative promoters are used for genes within maize chloroplast polycistronic transcription units. *Plant Cell*, **2**, 323-333.

Hanaoka, M., Kanamaru, K., Fujiwara, M., Takahashi, H. and Tanaka, K. (2005) Glutamyl-tRNA mediates a switch in RNA polymerase use during chloroplast biogenesis. *EMBO Rep*, **6**, 545-550.

Handa, H. (2008) Linear plasmids in plant mitochondria: peaceful coexistences or malicious invasions? *Mitochondrion*, **8**, 15-25.

Handa, H., Itani, K. and Sato, H. (2002) Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (Brassica napus L.). *Mol Genet Genomics*, **267**, 797-805.

Hara, K., Morita, M., Takahashi, R., Sugita, M., Kato, S. and Aoki, S. (2001) Characterization of two genes, Sig1 and Sig2, encoding distinct plastid sigma factors(1) in the moss Physcomitrella patens: phylogenetic relationships to plastid sigma factors in higher plants. *FEBS Lett*, **499**, 87-91.

Hayes, M.L., Reed, M.L., Hegeman, C.E. and Hanson, M.R. (2006) Sequence elements critical for efficient RNA editing of a tobacco chloroplast transcript in vivo and in vitro. *Nucleic Acids Res*, **34**, 3742-3754.

Hedtke, B. (1998) Isolation und Charakterisierung bakteriophagenähnlicher organellärer RNA-Polymerasen aus höheren Pflanzen: PhD Thesis, HU Berlin.

Hedtke, B., Börner, T. and Weihe, A. (1997) Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science*, **277**, 809-811.

Hedtke, B., Börner, T. and Weihe, A. (2000) One RNA polymerase serving two genomes. *EMBO Rep*, **1**, 435-440.

Hedtke, B., Legen, J., Weihe, A., Herrmann, R.G. and Börner, T. (2002) Six active phage-type RNA polymerase genes in Nicotiana tabacum. *Plant J*, **30**, 625-637.

Hedtke, B., Meixner, M., Gillandt, S., Richter, E., Börner, T. and Weihe, A. (1999a) Green fluorescent protein as a marker to investigate targeting of organellar RNA polymerases of higher plants in vivo. *Plant J*, **17**, 557-561.

Hedtke, B., Wagner, I., Börner, T. and Hess, W.R. (1999b) Inter-organellar crosstalk in higher plants: impaired chloroplast development affects mitochondrial gene and transcript levels. *Plant J*, **19**, 635-643.

Hegeman, C.E., Halter, C.P., Owens, T.G. and Hanson, M.R. (2005) Expression of complementary RNA from chloroplast transgenes affects editing efficiency of transgene and endogenous chloroplast transcripts. *Nucleic Acids Res*, **33**, 1454-1464.

Hellmund, D., Metzlaff, M. and Serfling, E. (1984) A transfer RNAArg gene of Pelargonium chloroplasts, but not a 5S RNA gene, is efficiently transcribed after injection into Xenopus oocyte nuclei. *Nucleic Acids Res*, **12**, 8253-8268.

Herrmann, J.M. (2003) Converting bacteria to organelles: evolution of mitochondrial protein sorting. *Trends Microbiol*, **11**, 74-79.

Herrmann, R.G., Bohnert, H.J., Kowallik, K.V. and Schmitt, J.M. (1975) Size, conformation and purity of chloroplast DNA of some higher plants. *Biochim Biophys Acta*, **378**, 305-317.

Hess, W.R. and Börner, T. (1999) Organellar RNA polymerases of higher plants. *Int Rev Cytol*, **190**, 1-59.

Hiesel, R., Wissinger, B., Schuster, W. and Brennicke, A. (1989) RNA editing in plant mitochondria. *Science*, **246**, 1632-1634.

Hohe, A., Egener, T., Lucht, J.M., Holtorf, H., Reinhard, C., Schween, G. and Reski, R. (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. *Curr Genet*, **44**, 339-347.

Holec, S., Lange, H., Canaday, J. and Gagliardi, D. (2008a) Coping with cryptic and defective transcripts in plant mitochondria. *Biochim Biophys Acta*, **1779**, 566-573.

Holec, S., Lange, H., Dietrich, A. and Gagliardi, D. (2008b) Polyadenylation-mediated RNA degradation in plant mitochondria. *Methods Enzymol*, **447**, 439-461.

Holec, S., Lange, H., Kühn, K., Alioua, M., Börner, T. and Gagliardi, D. (2006) Relaxed transcription in Arabidopsis mitochondria is counterbalanced by RNA stability control mediated by polyadenylation and polynucleotide phosphorylase. *Mol Cell Biol*, **26**, 2869-2876.

Hricova, A., Quesada, V. and Micol, J.L. (2006) The SCABRA3 nuclear gene encodes the plastid RpoTp RNA polymerase, which is required for chloroplast biogenesis and mesophyll cell proliferation in Arabidopsis. *Plant Physiol*, **141**, 942-956.

Hübschmann, T. and Börner, T. (1998) Characterisation of transcript initiation sites in ribosome-deficient barley plastids. *Plant Mol Biol*, **36**, 493-496.

Ichikawa, K., Sugita, M., Imaizumi, T., Wada, M. and Aoki, S. (2004) Differential expression on a daily basis of plastid sigma factor genes from the moss Physcomitrella patens. Regulatory interactions among PpSig5, the circadian clock, and blue light signaling mediated by cryptochromes. *Plant Physiol*, **136**, 4285-4298.

Ikeda, T.M. and Gray, M.W. (1999a) Characterization of a DNA-binding protein implicated in transcription in wheat mitochondria. *Mol Cell Biol*, **19**, 8113-8122.

Ikeda, T.M. and Gray, M.W. (1999b) Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. *Plant Mol Biol*, **40**, 567-578.

Iratni, R., Baeza, L., Andreeva, A., Mache, R. and Lerbs-Mache, S. (1994) Regulation of rDNA transcription in chloroplasts: promoter exclusion by constitutive repression. *Genes Dev*, **8**, 2928-2938.

Iratni, R., Diederich, L., Harrak, H., Bligny, M. and Lerbs-Mache, S. (1997) Organ-specific transcription of the rrn operon in spinach plastids. *J Biol Chem*, **272**, 13676-13682.

Iyer, L.M., Abhiman, S. and Aravind, L. (2008) A new family of polymerases related to superfamily A DNA polymerases and T7-like DNA-dependent RNA polymerases. *Biol Direct*, **3**, 39.

Jackson, C.J., Norman, J.E., Schnare, M.N., Gray, M.W., Keeling, P.J. and Waller, R.F. (2007) Broad genomic and transcriptional analysis reveals a highly derived genome in dinoflagellate mitochondria. *BMC Biol*, **5**, 41.

Jahn, D. (1992) Expression of the Chlamydomonas reinhardtii chloroplast tRNA(Glu) gene in a homologous in vitro transcription system is independent of upstream promoter elements. *Arch Biochem Biophys*, **298**, 505-513.

Jang, S.H. and Jaehning, J.A. (1991) The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial sigma factors. *J Biol Chem*, **266**, 22671-22677.

Jeruzalmi, D. and Steitz, T.A. (1998) Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J*, **17**, 4101-4113.

Kabeya, Y., Hashimoto, K. and Sato, N. (2002) Identification and characterization of two phage-type RNA polymerase cDNAs in the moss Physcomitrella patens: implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants. *Plant Cell Physiol*, **43**, 245-255.

Kabeya, Y., Kobayashi, Y., Suzuki, H., Itoh, J. and Sugita, M. (2007) Transcription of plastid genes is modulated by two nuclear-encoded alpha subunits of plastid RNA polymerase in the moss Physcomitrella patens. *Plant Journal*, **52**, 730-741.

Kabeya, Y. and Sato, N. (2005) Unique translation initiation at the second AUG codon determines mitochondrial localization of the phage-type RNA polymerases in the moss Physcomitrella patens. *Plant Physiol*, **138**, 369-382.

Kahlau and Bock (2008) Plastid Transcriptomics and Translatomics of Tomato Fruit Development and Chloroplast-to-Chromoplast Differentiation: Chromoplast Gene Expression Largely Serves the Production of a Single Protein. *The Plant Cell,* **20**, 856–874

Kanamaru, K. and Tanaka, K. (2004) Roles of chloroplast RNA polymerase sigma factors in chloroplast development and stress response in higher plants. *Biosci Biotechnol Biochem*, **68**, 2215-2223.

Kanki, T., Ohgaki, K., Gaspari, M., Gustafsson, C.M., Fukuoh, A., Sasaki, N., Hamasaki, N. and Kang, D. (2004) Architectural role of mitochondrial transcription factor A in maintenance of human mitochondrial DNA. *Mol Cell Biol*, **24**, 9823-9834.

Kapoor, S. and Sugiura, M. (1999) Identification of two essential sequence elements in the nonconsensus type II PatpB-290 plastid promoter by using plastid transcription extracts from cultured tobacco BY-2 cells. *Plant Cell*, **11**, 1799-1810.

Karpova, O.V., Kuzmin, E.V., Elthon, T.E. and Newton, K.J. (2002) Differential expression of alternative oxidase genes in maize mitochondrial mutants. *Plant Cell*, **14**, 3271-3284.

Klein, U., De Camp, J.D. and Bogorad, L. (1992) Two types of chloroplast gene promoters in Chlamydomonas reinhardtii. *Proc Natl Acad Sci U S A*, **89**, 3453-3457.

Kobayashi, Y., Dokiya, Y., Kumazawa, Y. and Sugita, M. (2002) Non-AUG translation initiation of mRNA encoding plastid-targeted phage-type RNA polymerase in Nicotiana sylvestris. *Biochem Biophys Res Commun*, **299**, 57-61.

Kobayashi, Y., Dokiya, Y. and Sugita, M. (2001a) Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts. *Biochem Biophys Res Commun*, **289**, 1106-1113.

Kobayashi, Y., Dokiya, Y., Sugiura, M., Niwa, Y. and Sugita, M. (2001b) Genomic organization and organ-specific expression of a nuclear gene encoding phage-type RNA polymerase in Nicotiana sylvestris. *Gene*, **279**, 33-40.

Kowallik, K.V. and Herrmann, R.G. (1972) Variable amounts of DNA related to the size of chloroplasts. IV. Three-dimensional arrangement of DNA in fully differentiated chloroplasts of Beta vulgaris L. *J Cell Sci*, **11**, 357-377.

Krause, K. (2008) From chloroplasts to "cryptic" plastids: evolution of plastid genomes in parasitic plants. *Curr Genet*, **54**, 111-121.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A. and Mikami, T. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Res*, **28**, 2571-2576.

Kühn, K. (2006) Analysis of components of the mitochondrial transcription machinery in Arabidopsis thaliana. PhD Thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I.

Kühn, K., Bohne, A.-V., Liere, K., Weihe, A. and Börner, T. (2007) Arabidopsis single-polypeptide RNA polymerases: accurate in vitro transcription of organellar genes. *Plant Cell*, **19**, 959–971.

Kühn, K., Richter, U., Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Börner, T., Millar, A.H., Small, I.D. and Whelan, J. (2009) Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. *Plant Cell*, **21**, 2762-2779.

Kühn, K., Weihe, A. and Börner, T. (2005) Multiple promoters are a common feature of mitochondrial genes in Arabidopsis. *Nucleic Acids Res*, **33**, 337-346.

Kusumi, Yara, Mitsui, Tozawa and Iba (2004) Characterization of a rice nuclear-encoded plastid RNA polymerase gene OsRpoTp. *Plant & cell physiology*, **45**, 1194-1201.

Kutschera, U. and Niklas, K.J. (2005) Endosymbiosis, cell evolution, and speciation. *Theory Biosci*, **124**, 1-24.

Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, **387**, 493-497.

Lange, B.M., Rujan, T., Martin, W. and Croteau, R. (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A*, **97**, 13172-13177.

Legen, Kemp, Krause, Profanter, Herrmann and Maier (2002) Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *The Plant journal : for cell and molecular biology*, **31**, 171-188.

Lerbs-Mache, S. (1993) The 110-kDa polypeptide of spinach plastid DNA-dependent RNA polymerase: single-subunit enzyme or catalytic core of multimeric enzyme complexes? *Proc Natl Acad Sci U S A*, **90**, 5509-5513.

Lerbs-Mache, S. (2000) Regulation of rDNA transcription in plastids of higher plants. *Biochimie*, **82**, 525-535.

Li, J., Maga, J.A., Cermakian, N., Cedergren, R. and Feagin, J.E. (2001) Identification and characterization of a Plasmodium falciparum RNA polymerase gene with similarity to mitochondrial RNA polymerases. *Mol Biochem Parasitol*, **113**, 261-269.

Li, X.Q., Zhang, M. and Brown, G.G. (1996) Cell-Specific Expression of Mitochondrial Transcripts in Maize Seedlings. *Plant Cell*, **8**, 1961-1975.

Liere, K. and Börner, T. (2006) Transcription of plastid genes. In *Regulation of Transcription in Plants.* (Grasser ed.), Blackwell Publishing Ltd., pp. 184-224.

Liere, K. and Börner, T. (2007) Transcription and transcriptional regulation in plastids. In *Topics in Current Genetics: Cell and Molecular Biology of Plastids.* (Bock, R. ed.) Berlin/Heidelberg: Springer, pp. 121-174.

Liere, K., Kaden, D., Maliga, P. and Börner, T. (2004) Overexpression of phage-type RNA polymerase RpoTp in tobacco demonstrates its role in chloroplast transcription by recognizing a distinct promoter type. *Nucl. Acids Res.*, **32**, 1159-1165.

Liere, K. and Link, G. (1994) Structure and expression characteristics of the chloroplast DNA region containing the split gene for tRNA(Gly) (UCC) from mustard (Sinapis alba L.). *Curr Genet*, **26**, 557-563.

Liere, K. and Maliga, P. (1999) In vitro characterization of the tobacco rpoB promoter reveals a core sequence motif conserved between phage-type plastid and plant mitochondrial promoters. *EMBO J*, **18**, 249-257.

Liere, K. and Maliga, P. (2001) Plastid RNA polymerases in higher plants. In *Regulatory aspects of photosynthesis* (Andersson, B. and Aro, E.-M. eds): Kluwer Academic Publishers, pp. 29-49.

Liere, K., Weihe, A. and Börner, T. (2011) The transcription machineries of plant mitochondria and chloroplasts: Composition, function, and regulation. *J Plant Physiol*, **168**, 1345–1360.

Lilly, J.W., Havey, M.J., Jackson, S.A. and Jiang, J. (2001) Cytogenomic analyses reveal the structural plasticity of the chloroplast genome in higher plants. *Plant Cell*, **13**, 245-254.

Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I., Town, C.D., Fujii, C.Y., Mason, T., Bowman, C.L., Barnstead, M., Feldblyum, T.V., Buell, C.R., Ketchum, K.A.,

Lee, J., Ronning, C.M., Koo, H.L., Moffat, K.S., Cronin, L.A., Shen, M., Pai, G., Van Aken, S., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carrera, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Copenhaver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C. (1999) Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. *Nature*, **402**, 761-768.

Link, G. (1996) Green life: Control of chloroplast gene transcription. *Bioessays*, **18**, 465-471.

Lohse, M., Drechsel, O. and Bock, R. (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*, **52**, 267-274.

Lonsdale, D.M. and Grienenberger, J.M. (1992) The mitochondrial genome of plants. In *In: Cell organelles.* (Herrmann Ed.), New York: Springer-Verlag, pp. 181–218.

Lupold, D.S., Caoile, A.G. and Stern, D.B. (1999) The maize mitochondrial cox2 gene has five promoters in two genomic regions, including a complex promoter consisting of seven overlapping units. *J Biol Chem*, **274**, 3897-3903.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.L., Mireau, H., Peeters, N., Renou, J.P., Szurek, B., Taconnat, L. and Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell*, **16**, 2089-2103.

Maier, R.M. and Schmitz-Linneweber, C. (2004) Plastid Genomes. In *Molecular Biology and Biotechnology of Plant Organelles*, pp. 115-150.

Maier, U.G., Bozarth, A., Funk, H.T., Zauner, S., Rensing, S.A., Schmitz-Linneweber, C., Börner, T. and Tillich, M. (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol*, **6**, 36.

Margulis, L. (1970) *Origin of Eukaryotic Cells*. New Haven, CT: Yale University Press.

Margulis, L. (1981) *Symbiosis in Cell Evolution.* San Francisco: Freeman.

Marienfeld, J., Unseld, M. and Brennicke, A. (1999) The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Trends Plant Sci*, **4**, 495-502.

Martin, W. and Herrmann, R.G. (1998) Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol*, **118**, 9-17.

Martin, W. and Müller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature*, **392**, 37–41.

Martin, W. and Schnarrenberger, C. (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr Genet*, **32**, 1-18.

Massey, S.E., Moura, G., Beltrao, P., Almeida, R., Garey, J.R., Tuite, M.F. and Santos, M.A. (2003) Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in Candida spp. *Genome Res*, **13**, 544-557.

Masters, B.S., Stohl, L.L. and Clayton, D.A. (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell*, **51**, 89-99.

Matsunaga, M. and Jaehning, J.A. (2004a) A mutation in the yeast mitochondrial core RNA polymerase, Rpo41, confers defects in both specificity factor interaction and promoter utilization. *J Biol Chem*, **279**, 2012-2019.

Matsunaga, M. and Jaehning, J.A. (2004b) Intrinsic promoter recognition by a "core" RNA polymerase. *J Biol Chem*, **279**, 44239-44242.

Matsushima, Y., Adan, C., Garesse, R. and Kaguni, L.S. (2005) Drosophila mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells. *J Biol Chem*, **280**, 16815-16820.

Matsushima, Y., Garesse, R. and Kaguni, L.S. (2004) Drosophila mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in schneider cells. *J Biol Chem*, **279**, 26900-26905.

McAllister, W.T. and Raskin, C.A. (1993) The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol Microbiol*, **10**, 1-6.

McCulloch, V., Seidel-Rogol, B.L. and Shadel, G.S. (2002) A human mitochondrial transcription factor is related to RNA adenine methyltransferases and binds S-adenosylmethionine. *Mol Cell Biol*, **22**, 1116-1125.

McDermott, P., Connolly, V. and Kavanagh, T.A. (2008) The mitochondrial genome of a cytoplasmic male sterile line of perennial ryegrass (Lolium perenne L.) contains an integrated linear plasmid-like element. *Theor Appl Genet*, **117**, 459-470.

Meinnel, T. and Giglione, C. (2008) Tools for analyzing and predicting N-terminal protein modifications. *Proteomics*, **8**, 626-649.

Metodiev, M.D., Lesko, N., Park, C.B., Camara, Y., Shi, Y., Wibom, R., Hultenby, K., Gustafsson, C.M. and Larsson, N.G. (2009) Methylation of 12S rRNA is necessary for in vivo stability of the small subunit of the mammalian mitochondrial ribosome. *Cell Metab*, **9**, 386-397.

Miranda, I., Silva, R. and Santos, M.A. (2006) Evolution of the genetic code in yeasts. *Yeast*, **23**, 203-213.

Miyagi, T., Kapoor, S., Sugita, M. and Sugiura, M. (1998) Transcript analysis of the tobacco plastid operon rps2/atpI/H/F/A reveals the existence of a non-consensus type II (NCII) promoter upstream of the atpI coding sequence. *Mol Gen Genet*, **257**, 299-307.

Murray, M.G. and Thompson, W.F. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*, **8**, 4321-4325.

Muse, S.V. (2000) Examining rates and patterns of nucleotide substitution in plants. *Plant Mol Biol*, **42**, 25-43.

Neuhaus, H. and Link, G. (1990) The chloroplast psbK operon from mustard (Sinapis alba L.): multiple transcripts during seedling development and evidence for divergent overlapping transcription. *Curr Genet*, **18**, 377-383.

Newton, K.J., Winberg, B., Yamato, K., Lupold, S. and Stern, D.B. (1995) Evidence for a novel mitochondrial promoter preceding the cox2 gene of perennial teosintes. *EMBO J*, **14**, 585-593.

Nickelsen, J. and Link, G. (1990) Nucleotide sequence of the mustard chloroplast genes trnH and rps19'. *Nucleic Acids Res*, **18**, 1051.

Nishimura, Y., Kikis, E.A., Zimmer, S.L., Komine, Y. and Stern, D.B. (2004) Antisense transcript and RNA processing alterations suppress instability of polyadenylated mRNA in chlamydomonas chloroplasts. *Plant Cell*, **16**, 2849-2869.

Noctor, G., Dutilleul, C., De Paepe, R. and Foyer, C.H. (2004) Use of mitochondrial electron transport mutants to evaluate the effects of redox state on photosynthesis,

stress tolerance and the integration of carbon/nitrogen metabolism. *J Exp Bot*, **55**, 49-57.

Noguchi, K. and Yoshida, K. (2008) Interaction between photosynthesis and respiration in illuminated leaves. *Mitochondrion*, **8**, 87-99.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A. and Kadowaki, K. (2002) The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics*, **268**, 434-445.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. and et al. (1992) Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol*, **223**, 1-7.

Odahara, M., Kuroiwa, H., Kuroiwa, T. and Sekine, Y. (2009) Suppression of repeat-mediated gross mitochondrial genome rearrangements by RecA in the moss Physcomitrella patens. *Plant Cell*, **21**, 1182-1194.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H. and Ozeki, H. (1986) Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature*, **322**, 572-574.

Oldenburg, D.J. and Bendich, A.J. (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *J Mol Biol*, **335**, 953-970.

Palmer, J.D. (1990) Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet*, **6**, 115-120.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L. and Song, K. (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci U S A*, **97**, 6960-6966.

Papadopoulos, J.S. and Agarwala, R. (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, **23**, 1073-1079.

Park, C.B., Asin-Cayuela, J., Camara, Y., Shi, Y., Pellegrini, M., Gaspari, M., Wibom, R., Hultenby, K., Erdjument-Bromage, H., Tempst, P., Falkenberg, M., Gustafsson, C.M. and Larsson, N.G. (2007) MTERF3 is a negative regulator of mammalian mtDNA transcription. *Cell*, **130**, 273-285.

Peeters, N., Chapron, A., Giritch, A., Grandjean, O., Lancelin, D., Lhomme, T., Vivrel, A. and Small, I. (2000) Duplication and quadruplication of Arabidopsis thaliana cysteinyl- and asparaginyl-tRNA synthetase genes of organellar origin. *J Mol Evol*, **50**, 413-423.

Peeters, N. and Small, I. (2001) Dual targeting to mitochondria and chloroplasts. *Biochim Biophys Acta*, **1541**, 54-63.

Perrin, R., Lange, H., Grienenberger, J.M. and Gagliardi, D. (2004) AtmtPNPase is required for multiple aspects of the 18S rRNA metabolism in Arabidopsis thaliana mitochondria. *Nucleic Acids Res*, **32**, 5174-5182.

Pesole, G., Gissi, C., De Chirico, A. and Saccone, C. (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol*, **48**, 427-434.

Pfannschmidt, T. and Link, G. (1997) The A and B forms of plastid DNA-dependent RNA polymerase from mustard (Sinapis alba L.) transcribe the same genes in a different developmental context. *Mol Gen Genet*, **257**, 35-44.

Prescott, E.M. and Proudfoot, N.J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A*, **99**, 8796-8801.

Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K. and Börner, T. (2010) Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *Plant J*, **64**, 948-959.

Pring, D., Brennicke, A. and Schuster, W. (1993) RNA editing gives a new meaning to the genetic information in mitochondria and chloroplasts. *Plant Mol Biol*, **21**, 1163-1170.

Rapp, W.D., Lupold, D.S., Mack, S. and Stern, D.B. (1993) Architecture of the maize mitochondrial atp1 promoter as determined by linker-scanning and point mutagenesis. *Mol Cell Biol*, **13**, 7232-7238.

Rapp, W.D. and Stern, D.B. (1992) A conserved 11 nucleotide sequence contains an essential promoter element of the maize mitochondrial atp1 gene. *EMBO J*, **11**, 1065-1073.

Rasmusson, A.G., Geisler, D.A. and Moller, I.M. (2008) The multiplicity of dehydrogenases in the electron transport chain of plant mitochondria. *Mitochondrion*, **8**, 47-60.

Reichheld, J.P., Khafif, M., Riondet, C., Droux, M., Bonnard, G. and Meyer, Y. (2007) Inactivation of thioredoxin reductases reveals a complex interplay between thioredoxin and glutathione pathways in Arabidopsis development. *Plant Cell*, **19**, 1851-1865.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin, I.T., Kuroki, Y., Toyoda, A., Suzuki, Y., Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W.B., Barker, E., Bennetzen, J.L., Blankenship, R., Cho, S.H., Dutcher, S.K., Estelle, M., Fawcett, J.A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K.A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D.R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P.J., Sanderfoot, A., Schween, G., Shiu, S.H., Stueber, K., Theodoulou, F.L., Tu, H., Van de Peer, Y., Verrier, P.J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A.C., Hasebe, M., Lucas, S., Mishler, B.D., Reski, R., Grigoriev, I.V., Quatrano, R.S. and Boore, J.L. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, **319**, 64-69.

Richter, U. (2003) Charakterisierung von zwei RpoT-Genen aus dem Laubmoos Physcomitrella patens. In *Abt. Genetik*: Humboldt University zu Berlin.

Richter, U., Kiessling, J., Hedtke, B., Decker, E., Reski, R., Börner, T. and Weihe, A. (2002) Two RpoT genes of Physcomitrella patens encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. *Gene*, **290**, 95-105.

Richter, U., Kühn, K., Okada, S., Brennicke, A., Weihe, A. and Börner, T. (2009) A mitochondrial rRNA dimethyladenosine methyltransferase in Arabidopsis. *Plant J*, **61**, 558-569.

Ringel, R., Sologub, M., Morozov, Y.I., Litonin, D., Cramer, P. and Temiakov, D. (2011) Structure of human mitochondrial RNA polymerase. *Nature*, **478**, 269-273.

Rodriguez-Ezpeleta, Brinkmann, Burger, Roger, Gray, Philippe and Lang (2007) Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Current Biology*, **17**, 1420-1425.

Rodriguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Loffelhardt, W., Bohnert, H.J., Philippe, H. and Lang, B.F. (2005) Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Current Biology*, **15**, 1325-1330.

Rogowska, A.T., Puchta, O., Czarnecka, A.M., Kaniak, A., Stepien, P.P. and Golik, P. (2006) Balance between transcription and RNA degradation is vital for Saccharomyces cerevisiae mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation. *Mol Biol Cell*, **17**, 1184-1193.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*, **4**, 406-425.

Sambrook, J., Russell, D.W. (2001) *Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell;* Cold Spring Harbor, N.Y. :: Cold Spring Harbor Laboratory.

Sanchez, H., Fester, T., Kloska, S., Schroder, W. and Schuster, W. (1996) Transfer of rps19 to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in Arabidopsis mitochondria. *EMBO J*, **15**, 2138-2149.

Scannell, D.R., Butler, G. and Wolfe, K.H. (2007) Yeast genome evolution--the origin of the species. *Yeast*, **24**, 929-942.

Schaefer, D.G. (2002) A new moss genetics: targeted mutagenesis in Physcomitrella patens. *Annu Rev Plant Biol*, **53**, 477-501.

Schaefer, D.G. and Zryd, J.P. (1997) Efficient gene targeting in the moss Physcomitrella patens. *Plant J*, **11**, 1195-1206.

Schinkel, A.H., Groot Koerkamp, M.J., Stuiver, M.H., Van der Horst, G.T. and Tabak, H.F. (1987) Effect of point mutations on in vitro transcription from the promoter for the large ribosomal RNA gene of yeast mitochondria. *Nucleic Acids Res*, **15**, 5597-5612.

Schnable, P.S. and Wise, R.P. (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. **3**, 175-180.

Schnarrenberger, C. and Martin, W. (2002) Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur J Biochem*, **269**, 868-883.

Schuster, W. (1993) Ribosomal protein gene rpl5 is cotranscribed with the nad3 gene in Oenothera mitochondria. *Mol Gen Genet*, **240**, 445-449.

Schweer, J., Loschelder, H. and Link, G. (2006) A promoter switch that can rescue a plant sigma factor mutant. *FEBS Lett*, **580**, 6617-6622.

Seidel-Rogol, B.L., McCulloch, V. and Shadel, G.S. (2003) Human mitochondrial transcription factor B1 methylates ribosomal RNA at a conserved stem-loop. *Nat Genet*, **33**, 23-24.

Serino, G. and Maliga (1998) RNA polymerase subunits encoded by the plastid rpo genes are not shared with the nucleus-encoded plastid enzyme. *Plant Physiology*, **117**, 1165-1170.

Sharwood, R.E., Hotto, A.M., Bollenbach, T.J. and Stern, D.B. (2009) Overaccumulation of the chloroplast antisense RNA AS5 is correlated with decreased abundance of 5S rRNA in vivo and inefficient 5S rRNA maturation in vitro. *RNA*, **17**, 230-243.

Shutt, T.E. and Gray, M.W. (2006a) Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends Genet*, **22**, 90-95.

Shutt, T.E. and Gray, M.W. (2006b) Homologs of mitochondrial transcription factor B, sparsely distributed within the eukaryotic radiation, are likely derived from the dimethlyadenosine methyltransferase of the mitochondrial endosymbiont. . *Mol Biol Evol*, **23**, 1169–1179.

Shutt, T.E., Lodeiro, M.F., Cotney, J., Cameron, C.E. and Shadel, G.S. (2010) Core human mitochondrial transcription apparatus is a regulated two-component system in vitro. *Proc Natl Acad Sci U S A*, **107**, 12133-12138.

Silhavy, D. and Maliga, P. (1998) Mapping of promoters for the nucleus-encoded plastid RNA polymerase (NEP) in the iojap maize mutant. *Curr Genet*, **33**, 340-344.

Simpson, Inagaki and Roger (2006) Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. *Molecular Biology and Evolution*, **23**, 615-625.

Small, I.D. and Peeters, N. (2000) The PPR motif - a TPR-related motif prevalent in plant organellar proteins. *Trends Biochem Sci*, **25**, 46-47.

Sousa, R. (1996) Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem Sci*, **21**, 186-190.

Sousa, R., Chung, Y.J., Rose, J.P. and Wang, B.C. (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature*, **364**, 593-599.

Sriraman, P., Silhavy, D. and Maliga, P. (1998a) The phage-type PclpP-53 plastid promoter comprises sequences downstream of the transcription initiation site. *Nucleic Acids Res*, **26**, 4874-4879.

Sriraman, P., Silhavy, D. and Maliga, P. (1998b) Transcription from heterologous rRNA operon promoters in chloroplasts reveals requirement for specific activating factors. *Plant Physiol*, **117**, 1495-1499.

Stegemann, S. and Bock, R. (2006) Experimental reconstruction of functional gene transfer from the tobacco plastid genome to the nucleus. *Plant Cell*, **18**, 2869-2878.

Steitz, T.A. (2004) The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase. *Curr Opin Struct Biol*, **14**, 4-9.

Strepp, R., Scholz, S., Kruse, S., Speth, V. and Reski, R. (1998) Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc Natl Acad Sci U S A*, **95**, 4368-4373.

Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R. and Jiang, J. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci U S A*, **98**, 5099-5103.

Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A. and Sugiura, M. (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol Genet Genomics*, **272**, 603-615.

Suzuki, T., Nakamura, C., Mori, N. and Kaneda, C. (1995) Overexpression of mitochondrial genes in alloplasmic common wheat with a cytoplasm of wheatgrass (Agropyron trichophorum) showing depressed vigor and male sterility. *Plant Molecular Biology*, **27**, 553-565.

Swiatecka-Hagenbruch, M., Emanuel, C., Hedtke, B., Liere, K. and Börner, T. (2008) Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Res*, **36**, 785-792.

Swiatecka-Hagenbruch, M., Liere, K. and Börner, T. (2007) High diversity of plastidial promoters in Arabidopsis thaliana. *Mol Genet Genomics*, **277**, 725-734.

Tan, X.Y., Liu, X.L., Wang, W., Jia, D.J., Chen, L.Q., Zhang, X.Q. and Ye, D. (2010) Mutations in the Arabidopsis nuclear-encoded mitochondrial phage-type RNA polymerase gene RPOTm led to defects in pollen tube growth, female gametogenesis and embryogenesis. *Plant Cell Physiol*, **51**, 635-649.

Tandara, H. (2000) *In vitro*-Transkriptions- und *in vivo*-Importstudien zur Charakterisierung von Transkriptionsfaktoren pflanzlicher Organellen: Diploma thesis, HU Berlin.

Tang, G.Q., Bandwar, R.P. and Patel, S.S. (2005) Extended upstream A-T sequence increases T7 promoter strength. *J Biol Chem*, **280**, 40707-40713.

Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M. and Sato, N. (2007) The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. *Mol Biol Evol*, **24**, 699-709.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**, 796-815.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**, 4673-4680.

Timmis J.N., A.M.A., Huang C.Y., Martin W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics*, **5**, 123–135.

Topping, J. and Leaver, C. (1990) Mitochondrial gene expression during wheat leaf development. *Planta*, **182**, 399-407.

Torres, T.T., Dolezal, M., Schlotterer, C. and Ottenwalder, B. (2009) Expression profiling of Drosophila mitochondrial genes via deep mRNA sequencing. *Nucleic Acids Res*, **37**, 7509-7518.

Tovar, J., Leon-Avila, G., Sanchez, L.B., Sutak, R., Tachezy, J., van der Giezen, M., Hernandez, M., Muller, M. and Lucocq, J.M. (2003) Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation. *Nature*, **426**, 172-176.

Tracy, R.L. and Stern, D.B. (1995) Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr Genet*, **28**, 205-216.

Unseld, M., Marienfeld, J.R., Brandt, P. and Brennicke, A. (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nat Genet*, **15**, 57-61.

Van Aken, O., Giraud, E., Clifton, R. and Whelan, J. (2009) Alternative oxidase: a target and regulator of stress responses. *Physiol Plant*, **137**, 354-361.

Van Aken, O., Pecenkova, T., van de Cotte, B., De Rycke, R., Eeckhout, D., Fromm, H., De Jaeger, G., Witters, E., Beemster, G.T., Inze, D. and Van Breusegem, F. (2007) Mitochondrial type-I prohibitins of Arabidopsis thaliana are required for supporting proficient meristem development. *Plant J*, **52**, 850-864.

Wall, M.K., Mitchenall, L.A. and Maxwell, A. (2004) Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A*, **101**, 7821-7826.

Weihe, A. (2004) The transcription of plant organelle genomes. In *Molecular biology and biotechnology of plant organelles* (Daniell, H. and Chase, C.D. eds): Springer, 213-237.

Weihe, A., Hedtke, B. and Börner, T. (1997) Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant Chenopodium album. *Nucleic Acids Res*, **25**, 2319-2325.

Wettstein-Edwards, J., Ticho, B.S., Martin, N.C., Najarian, D. and Getz, G.S. (1986) In vitro transcription and promoter strength analysis of five mitochondrial tRNA promoters in yeast. *J Biol Chem*, **261**, 2905-2911.

Williams, K.P., Sobral, B.W. and Dickerman, A.W. (2007) A robust species tree for the alphaproteobacteria. *J Bacteriol*, **189**, 4578-4586.

Winkley, C.S., Keller, M.J. and Jaehning, J.A. (1985) A multicomponent mitochondrial RNA polymerase from Saccharomyces cerevisiae. *J Biol Chem*, **260**, 14214-14223.

Wischmann, C. and Schuster, W. (1995) Transfer of rps10 from the mitochondrion to the nucleus in Arabidopsis thaliana: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett*, **374**, 152-156.

Wolfe, K.H., Li, W.H. and Sharp, P.M. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*, **84**, 9054-9058.

Wu, C.Y., Lin, C.H. and Chen, L.J. (1997) Identification of the transcription start site for the spinach chloroplast serine tRNA gene. *FEBS Lett*, **418**, 157-161.

Xie, G. and Allison, L.A. (2002) Sequences upstream of the YRTA core region are essential for transcription of the tobacco atpB NEP promoter in chloroplasts in vivo. *Curr Genet*, **41**, 176-182.

Yamasaki, S., Konno, N. and Kishitani, S. (2004) Overexpression of mitochondrial genes is caused by interactions between the nucleus of Brassica rapa and the cytoplasm of Diplotaxis muralis in the leaves of alloplasmic lines of B. rapa. *J Plant Res*, **117**, 339-344.

Yin, C., Richter, U., Börner, T. and Weihe, A. (2009) Evolution of phage-type RNA polymerases in higher plants: characterization of the single phage-type RNA polymerase gene from Selaginella moellendorffii. *J Mol Evol*, **68**, 528-538.

Yin, C., Richter, U., Börner, T. and Weihe, A. (2010) Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm Nuphar advena encodes two mitochondrial and one plastid phage-type RNA polymerases. *BMC Evol Biol*, **10**, 379.

Yoo, S.D., Cho, Y.H. and Sheen, J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc*, **2**, 1565-1572.

Yoshida, K., Watanabe, C., Kato, Y., Sakamoto, W. and Noguchi, K. (2008) Influence of chloroplastic photo-oxidative stress on mitochondrial alternative oxidase capacity and respiratory properties: a case study with Arabidopsis yellow variegated 2. *Plant Cell Physiol*, **49**, 592-603.

Young, D.A. and Lonsdale, D.M. (1997) Evidence that plant mitochondrial transcription requires promoter-specific factors. *Maize Genet Coop Newsletter*, **71**, 69-71.

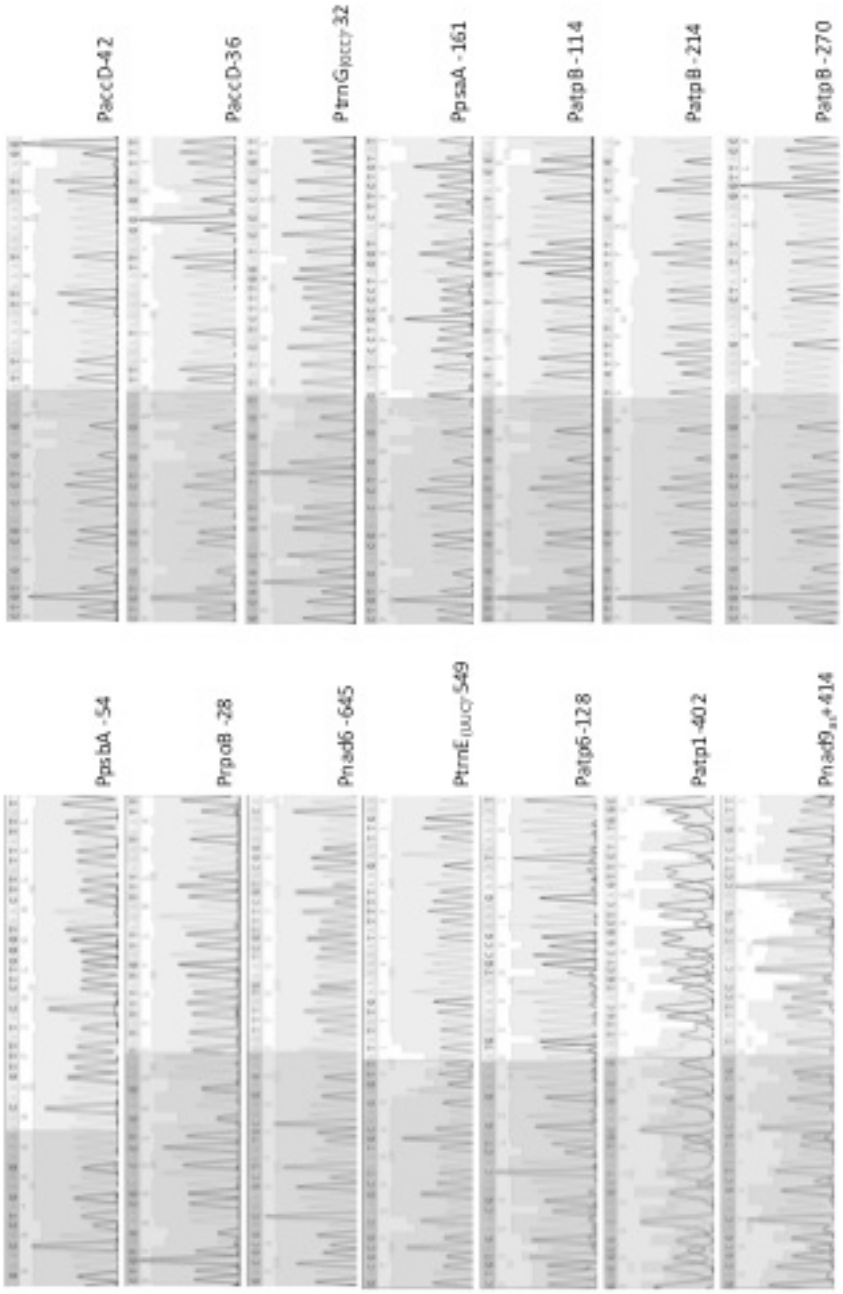
Zaegel, V., Guermann, B., Le Ret, M., Andres, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J.M. and Imbault, P. (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in Arabidopsis. *Plant Cell*, **18**, 3548-3563.

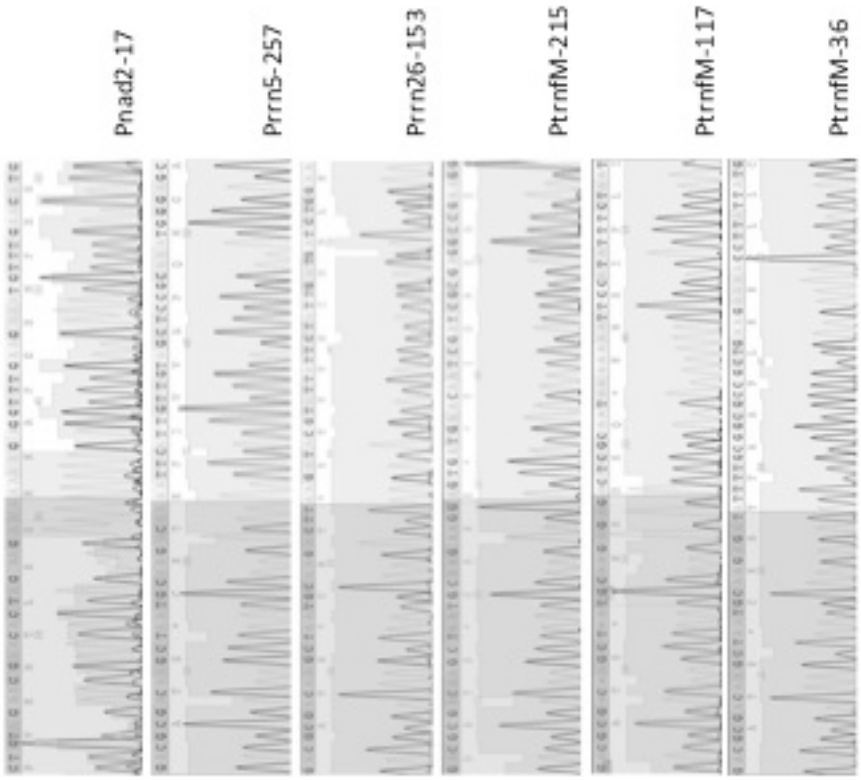
Zsigmond, L., Rigo, G., Szarka, A., Szekely, G., Otvos, K., Darula, Z., Medzihradszky, K.F., Koncz, C., Koncz, Z. and Szabados, L. (2008) Arabidopsis PPR40 connects abiotic stress responses to mitochondrial electron transport. *Plant Physiol*, **146**, 1721-1737.

6 Annex

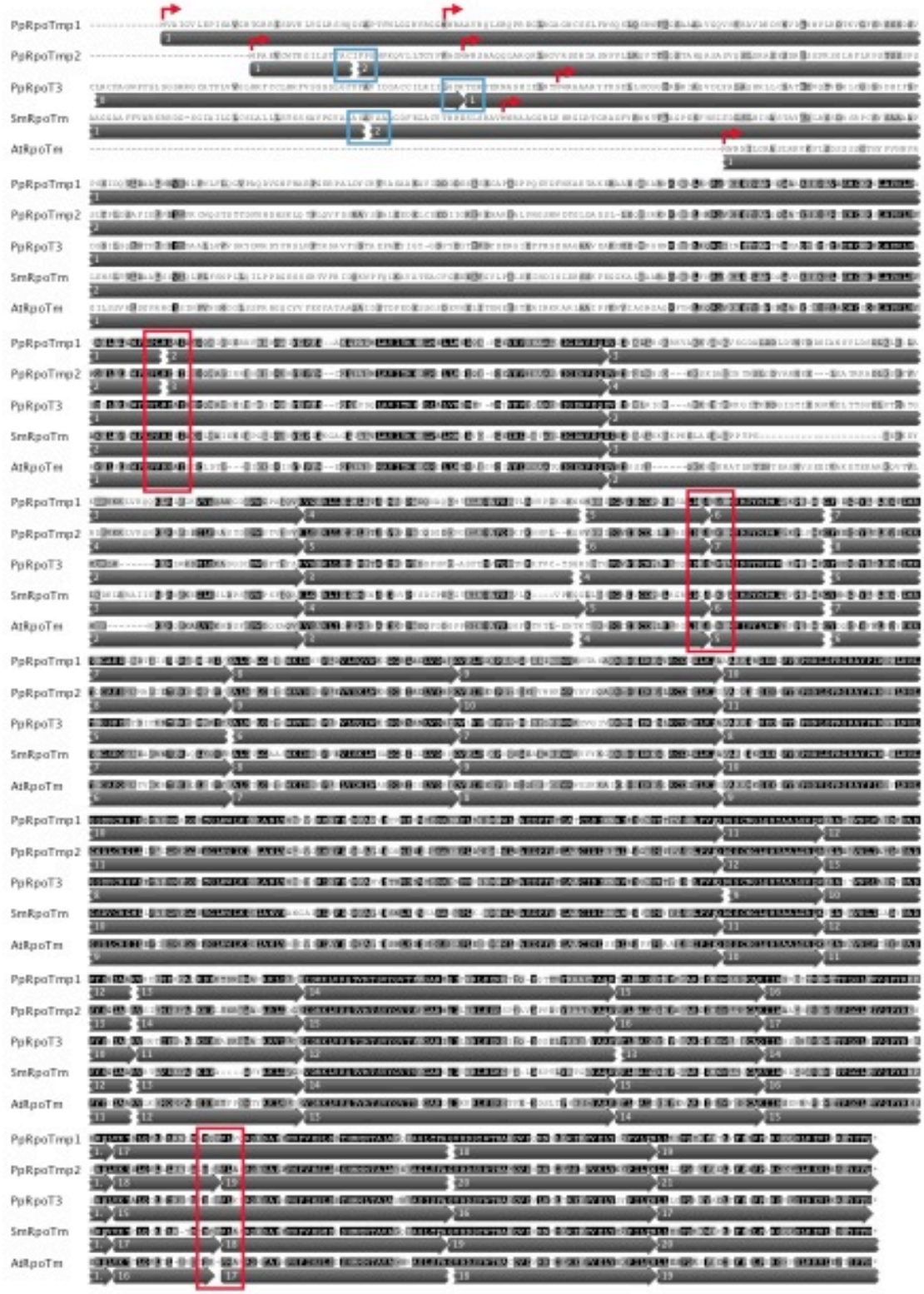
6.1 Supplementary Figures and Tables

Supplementary Figure 1: Sequences surrounding transcription initiation sites detected in this study. Initiating nucleotides are in bold; observed promoter cores and the frequent GAA motif are highlighted. Organellar promoter regions as deduced from the identified TSS in *Physcomitrella patens* were aligned to known NEP and PEP promoter regions from *Nicotiana tabacum* and mitochondrial promoters from *Arabidopsis thaliana*.

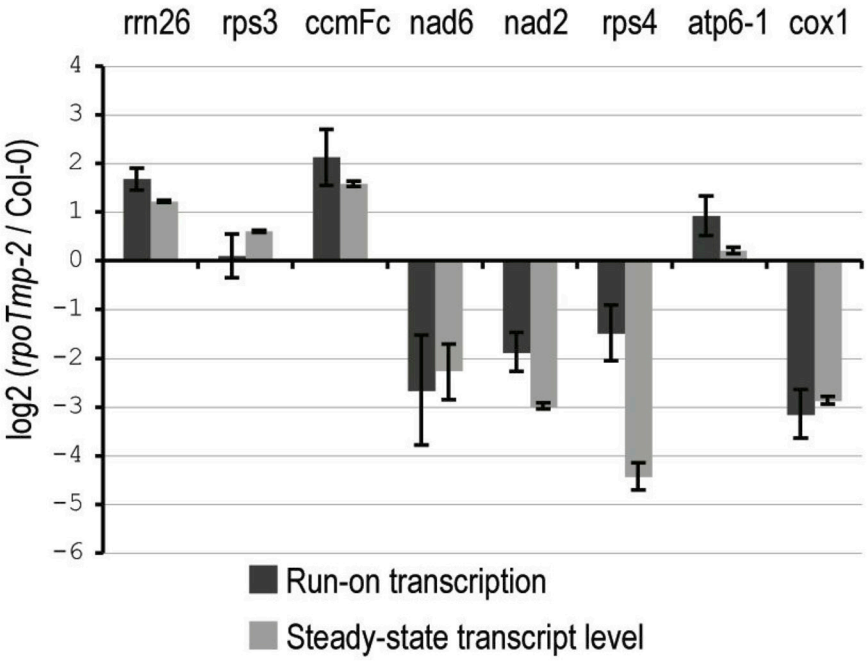

PpsbA -54
PrpoB -28
Pnad6 -645
PtrnE(UUC)-549
Patp6-128
Patp1-402
Pnad9 +414
PaccD-42
PaccD-36
PtrnG(GCC)-32
PpsaA -161
PatpB -114
PatpB -214
PatpB -270



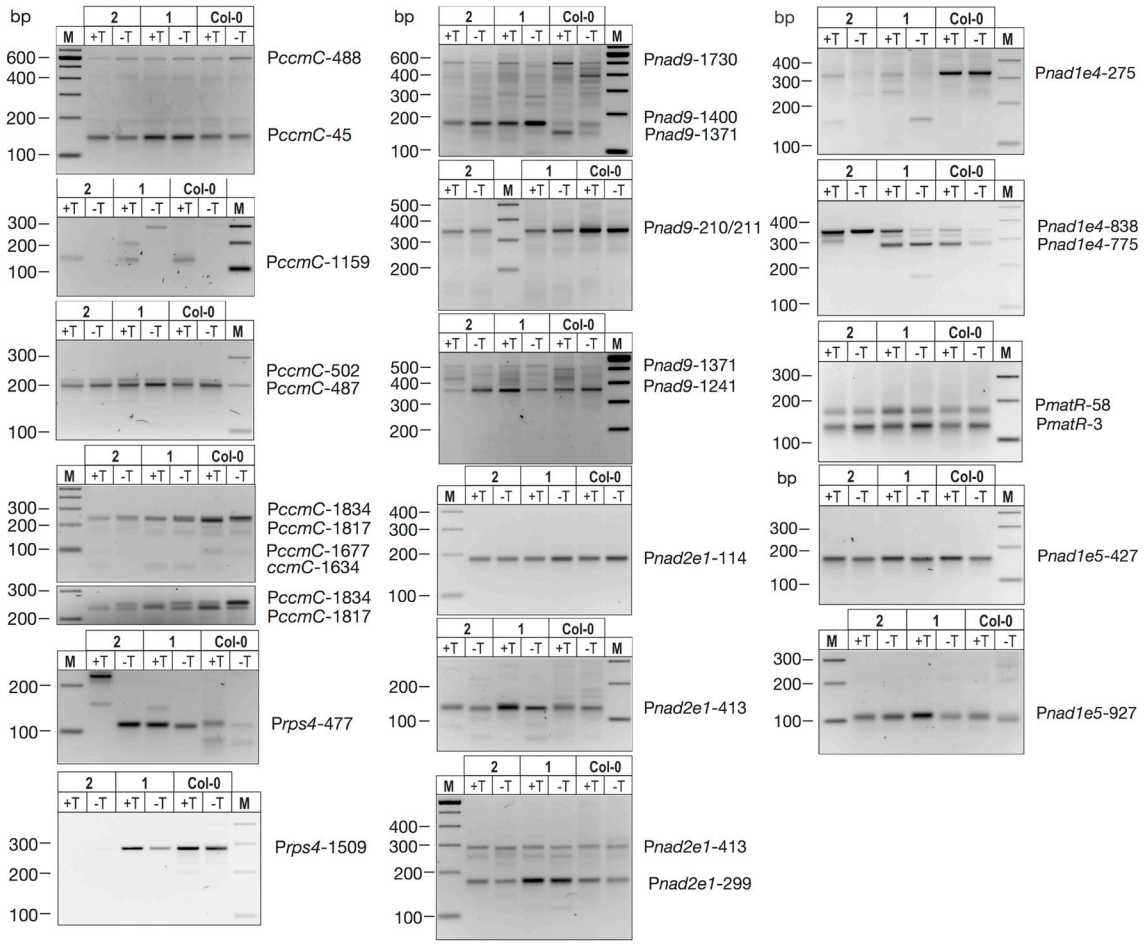
Supplementary Figure 2: Sequencing of 5'-RACE products across ligation sites. Chromatograms show the sequences at ligation sites of cloned 5'-RACE products derived from transcripts initiated at the indicated promoters. The portion of the RNA linker is indicated by highlighting in dark grey.



Supplementary Figure 3: ClustalW derived protein alignment. All three RpoT proteins from *Physcomitrella* were aligned with SmRpoTm as of *Selaginella moellendorfii* and AtRpoTm as of *Arabidopsis thaliana* using the ClustalW algorithm. Grey arrows below the sequence indicate exons and numbers inside indicate exon numbers. While red arrows mark putative translational start sites, blue boxes highlight intronic insertions in the putative 5'UTRs. Red boxes mark intron positions found in at least one of the *Physcomitrella RpoT* genes, but lacking in *PpRpoT3*.

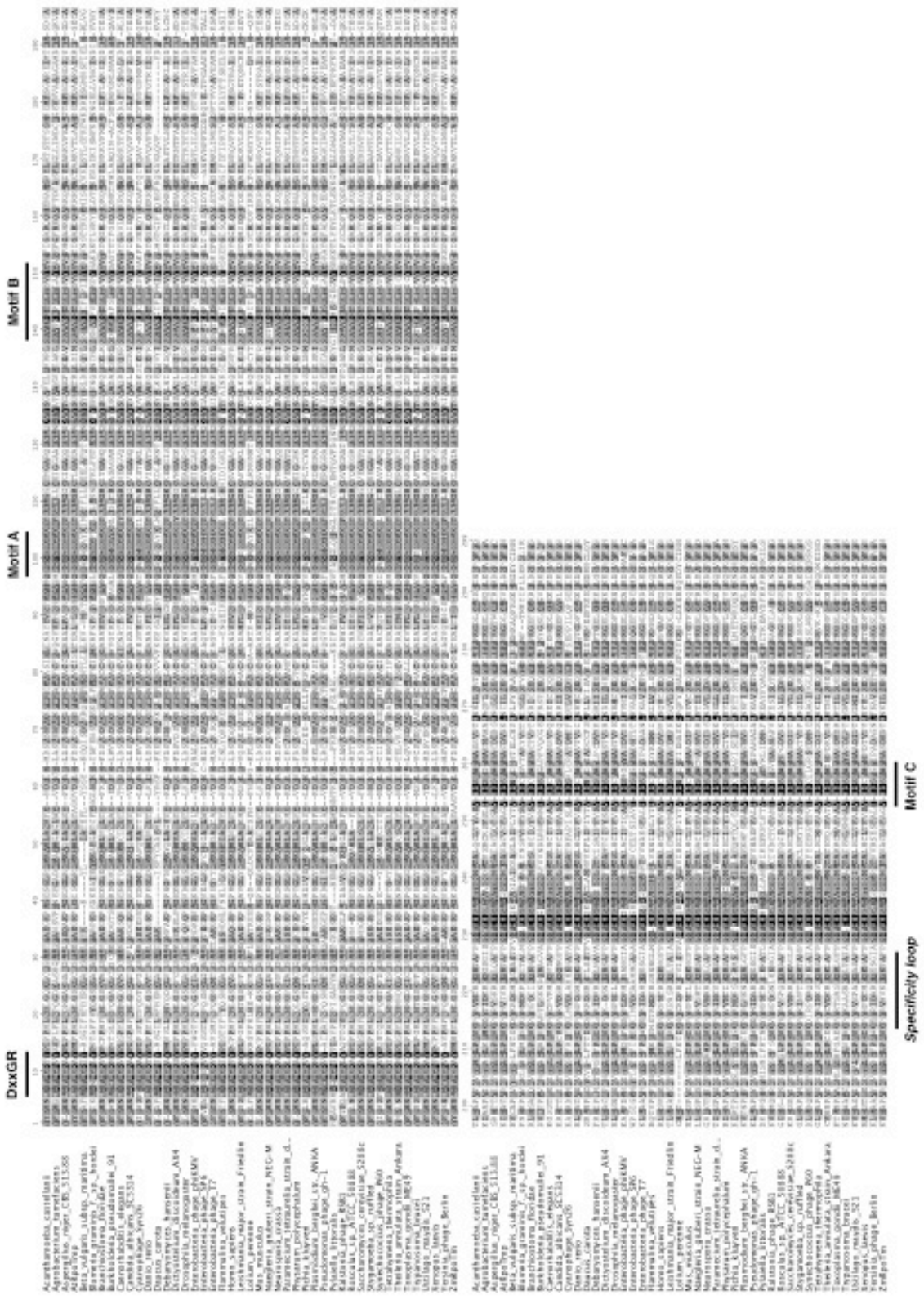


Supplementary Figure 4: Run-On transcription assay. Mitochondrial run-on transcription rates determined by Dr. Kristina Kühn for rpoTmp-2 (dark gray) and steady-state transcript levels measured for rpoTmp-2 rosettes (light gray). Taken from Kühn et al. (2009).

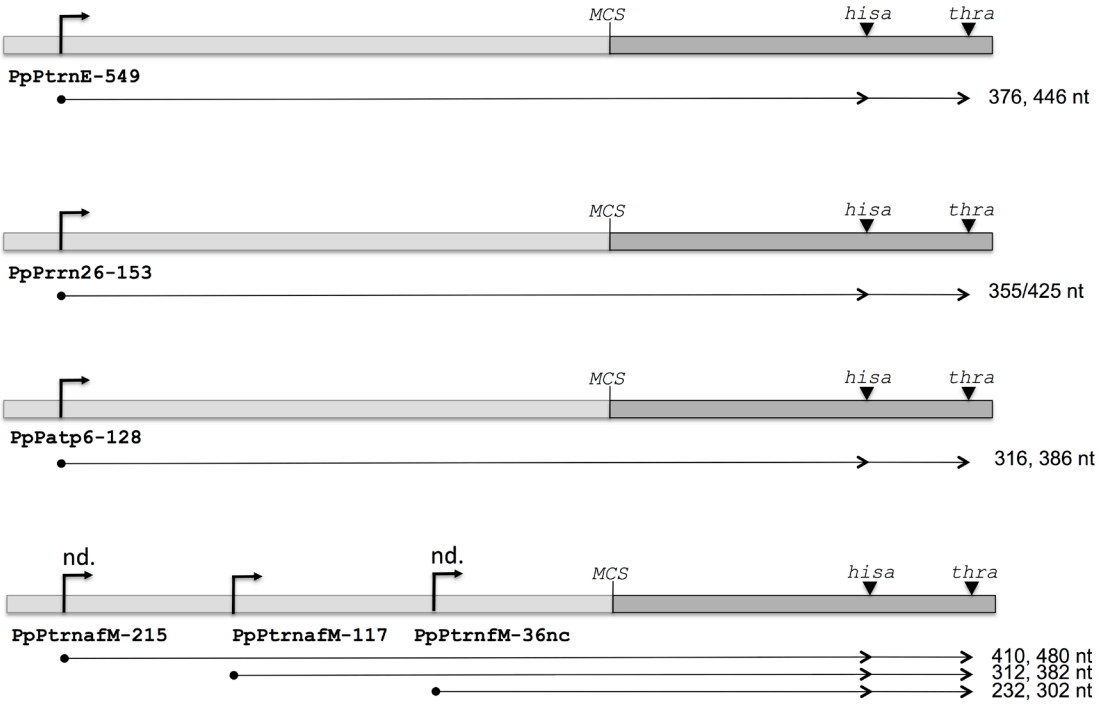


Supplementary Figure 5: 5'-RACE of ccmC, rps4, nad9, nad2e1, nad1e4, matR and nad1e5 transcripts. 5'-RACE analysis was performed as detailed before. PCR products from TAP-treated (+T) and untreated RNA (-T) were separated on agarose gels alongside a molecular weight marker (M). Transcript 5' termini were analysed through cloning and sequencing of marked 5'-RACE products. Products from transcript 5' ends are labelled with the name of the corresponding promoter or processing site as listed in Table 17, Figure 27 and 29.
Sequencing of ccmC transcripts mapping to position -1834 was insufficient to support these transcripts being primary (putative TSS). No major TAP-specific 5' end was found for PccmC-487, which showed a variety of ends including a 5' end reported by Forner et al. (2007). The CGTA tetranucleotide motive might suggest PccmC-487 functioning as TSS, but experimental proof is lacking. While no differences were seen in PccmC-45 5'-RACE gel patterns between +TAP and –TAP samples, sequencing identified a 5' end that was exclusive to TAP-treated RNA and therefore represents a primary end. Sequencing of products mapping to PccmC-1677 and PccmC-1817 were found to be TAP-specific, and therefore primary, transcript 5' ends. No products mapping to P*ccmC*-1159 were amplified from –TAP samples, identifying this site as a TSS.
In a first 5'-RACE assay targeting rps4, no specific rpoTmp-2 products were amplified from TAP-treated RNA. When the whole 5'-RACE was replicated, specific products were seen for rpoTmp-2 but not rpoTmp-1. This is seemingly due to the low abundance of 5' ends mapping to Prps4-477 and Prps4-1509 in mutants effecting the reproducibility of the experiment likely at the stage of reverse transcription. For Prps4-477, slower migration in +T lanes is observed. This is a typical observation for primary transcripts that are post-transcriptionally processed at their 5' ends by only a few nucleotides. For nad9 transcript 5' ends only Pnad9-1371 shows a TAP-specific product in 5'-RACE and is therefore a verified promoter. All others are considered as putative promoters since evidence identifying these sites as TSS is insufficient. For Pnad2e1-299 and Pnad2e1-413, slower migration in +T lanes is observed (processed by only a few nucleotides). Pnad2e1-114 is considered as a putative promoter since evidence identifying this site as a TSS is insufficient. No signal is seen in the displayed 5'-RACE experiments for Pnad1e4-775 in rpoTmp-2. However in replicates of this experiment the signal has been

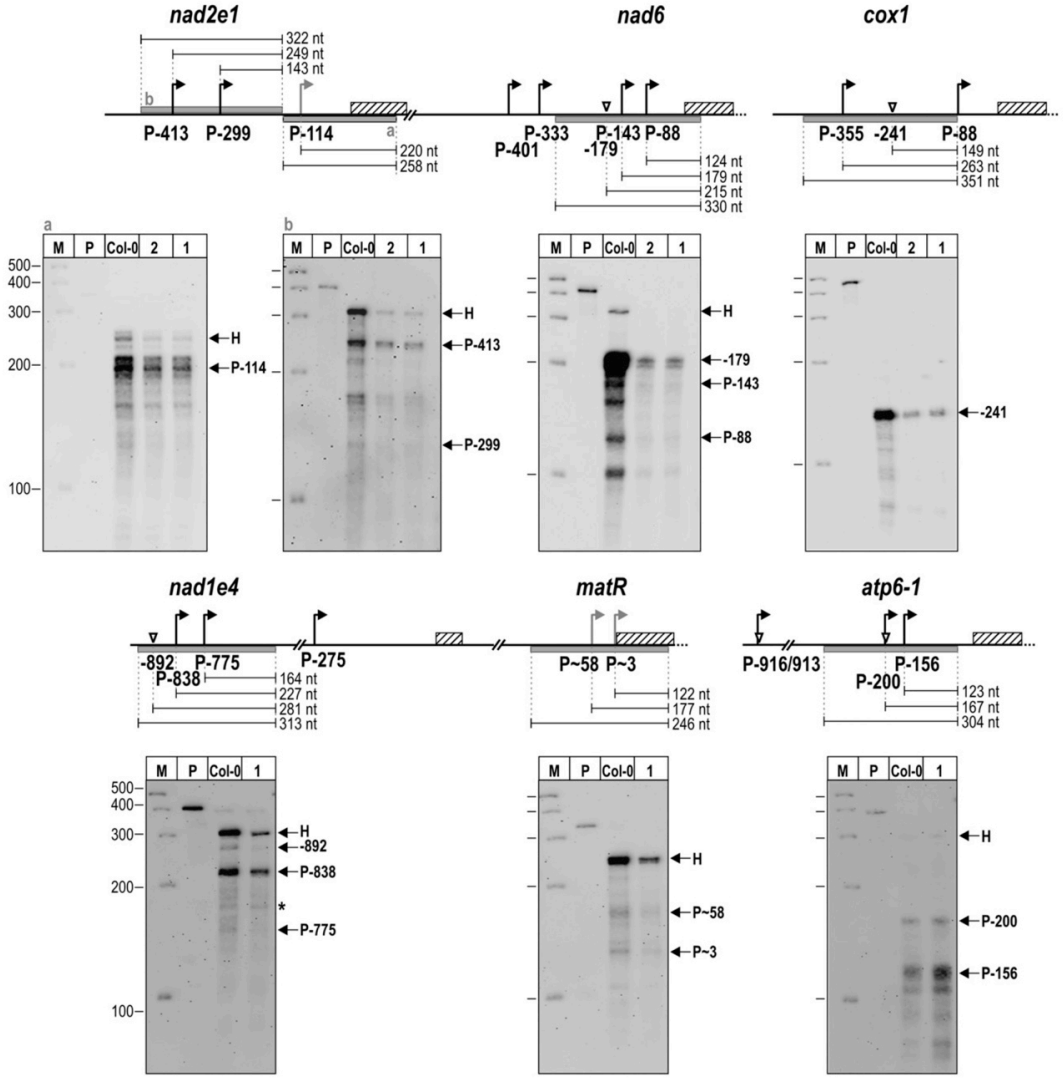
detected (data not shown). The *matR* 5’ ends were highly diverse and were inconclusive when comparing TAP-treated and untreated samples in gel images. The TAP-specific 5’ end detected mapping to position -3 in a highly A-rich sequence stretch can not be ascribed to a transcription initiation site with complete certainty. Even though no transcripts mapping to P*matR*-58 were analysed from TAP untreated RNA, this site is considered a putative promoter because of its ATTA element and the generally A/T-rich sequence (see Tab. 17 and Fig. 29). For *nad1e5* products from TAP treated transcripts are migrating more slowly than untreated (–T) (processed by only a few nucleotides).



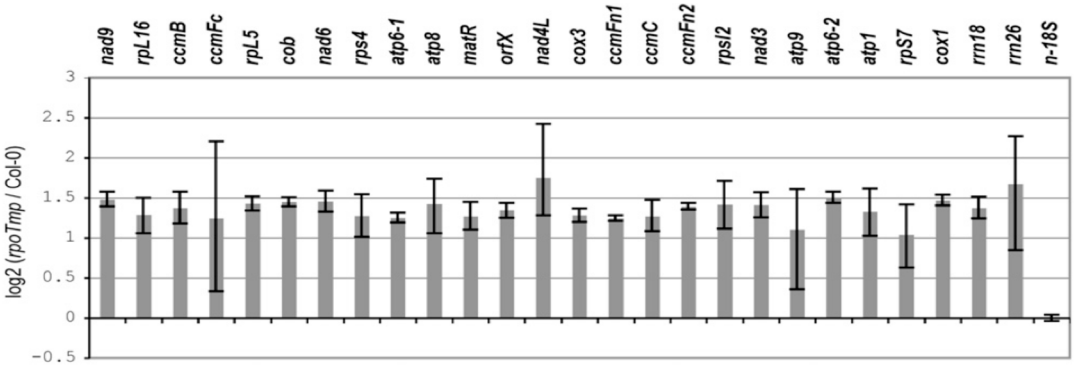
Supplementary Figure 6: Multiple alignment of monomeric DNA dependent RNAPs. Only a representative subset of all the sequences used for the phylogenetic analysis is shown here, including phage polymerases, bacterial ORF derived polypeptides and eukaryotic nucleus and mitochondrial plastid encoded phage-type RNAPs. The initially generated COBALT alignment (standard option settings) was subsequently imported into GENEIOUS v4.5 (Drummond *et al.* 2008) and further modified. Specifically, poorly aligned regions (mainly due to the higher divergency of the mitochondrial plasmid-encoded polymerases) were excluded from the alignment.



Supplementary Figure 7: Generation of template plasmids for *in vitro* transcription. pKL23 template plasmids were generated as described for pKL-atp8 (see Fig.17). Run-of products expected from initiation at PpPtrnE-549, PpPrrn26-153, PpPatp6-128 and PpPtrnfM-36nc, PpPtrnfM-117, PpPtrnfM-215 and termination at *hisa* and *thra* are indicated by horizontal black arrows labelled with the respective RNA length. While black triangles mark the positions of terminator structures on the vector backbone (dark grey), mtDNA inserted into pKL23 is shown in pale grey.



Supplementary Figure 8: Transcript 5' end abundances in rpoTmp and Wild-Type Seedlings.
Figure taken from Kühn et al. (2009). Rnase Protection Assays (RPA) performed on total seedling RNA from rpoTmp-1 (lanes labeled 1), rpoTmp-2 (2), and the wild type (Col-0). Protected RNA fragments were separated alongside a molecular weight marker (lane M; sizes in nucleotides to the left) and the probe (P) used. Arrows mark specific protected fragments, which are labeled with the corresponding name of the promoter or processing site. Signals from completely protected homologous segments of the probe are marked H.
5'-untranslated regions of *nad2e1*, *nad6*, *cox1*, *nad1e4*, *matR*, and *atp6-1* are illustrated above the gel images, indicating positions of promoters by bent arrows, processing sites by triangles; coding sequences by hatched bars. Gray bars below mark homologous segments of the riboprobes (complementary). Nucleotide sizes indicate the expected fragment length.



Supplementary Figure 9: Mitochondrial gene copy number in *AtrpoTmp* over Col-0 plants.
Figure taken from Kühn et al. (2009). Gene copy numbers are shown as the log2 ratios of DNA levels in mutants compared with levels in wild-type plants (Col-0).

Supplementary Table 1: List of monomeric RNAP proteins included in the cross-kingdom phylogenetic analysis. Accession numbers and taxon names are listed, together with abbreviations used for the labeling of the plant RpoT proteins (m mitochondrial, p plastid, mp dual).

Abbr.	Taxon/Name/Homolog	Annotation
	Acanthamoeba castellanii	ABD59451
	Acyrthosiphon pisum	XP001946877
	Aedes aegypti	XP001654145
	Agaricus bitorquis	P33539
	Agrobacterium tumefaciens str. C58	NP354202
	Ailuropoda melanoleuca	XP002923551
	Ajellomyces capsulatus G186AR	EEH06824
AlRpoTm	Arabidopsis lyrata RpoTm	EFH64954
AlRpoTmp	Arabidopsis lyrata RpoTmp	EFH49996
AlRpoTp	Arabidopsis lyrata RpoTp	EFH56807
	Arthroderma otae	XP002844119
	Ashbya gossypii ATCC 10895	NP983626
	Aspergillus niger CBS 513.88	XP001397108
AtRpoTm	Arabidopsis thaliana RpoTm	P92969
AtRpoTmp	Arabidopsis thaliana RpoTmp	CAC17120
AtRpoTp	Arabidopsis thaliana RpoTp	O24600
	Azorhizobium caulinodans ORS 571	YP001526523
	Babesia bovis T2Bo	XP001611431
BdRpoTm	Brachypodium distachyon RpoTm	Bradi4g08880 (http://www.phytozome.com)
BdRpoTp	Brachypodium distachyon RpoTp	Bradi1g31070 (http://www.phytozome.com)
	Beta vulgaris subsp. maritima	CAA71809
	Blastocystis hominis	CBK19597
	Blumeria graminis f. sp. hordei	NP862205
BoRpoTm	Brassica oleracea RpoTm	XP_002308414.1
	Bos taurus	XP580925
	Botryotinia fuckeliana B05.10	XP001545917
	Branchiostoma floridae	XP002610004
	Brassica napus	NP862328
	Brugia malayi	XP001896353
	Burkholderia oklahomensis EO147	ZP02357903
	Burkholderia pseudomallei 91	ZP02453457
	Burkholderia pseudomallei 1710b	YP333056
	Burkholderia thailandensis MSMB43	ZP02468154
	Caenorhabditis briggsae AF16	CAP36930
	Caenorhabditis elegans	NP001122532
	Camponotus floridanus	EFN75004
	Candida albicans SC5314	EAL02001.1
	Candida albicans SC5314	XP719003
	Candida albicans WO-1	EEQ47012.1
	Candida dubliniensis CD36	CAX40751.1
	Candida tropicalis MYA-3404	EER31317.1
	Candidatus Puniceispirillum marinum IMCC1322	YP003550722
CaRpoTm	Chenopodium album RpoTm	CAA69305
	Chaetomium globosum CBS 148.51	XP001223554
	Ciona intestinalis	XP002124660
	Claviceps purpurea	P22372
	Clavispora lusitaniae	XP002615589
	Coccidioides immitis RS	XP001246500
	Cryptococcus neoformans	XP570484
CsRpoTmp	Cleome spinosa RpoTmp	DQ415921
	Culex quinquefasciatus	XP001842642
CusRpoTm	Cucumis sativus RpoTm	Cucsa.284310 (http://www.phytozome.com)
CusRpoTmp	Cucumis sativus RpoTmp	Cucsa.238450 (http://www.phytozome.com)
CusRpoTp	Cucumis sativus RpoTp	Cucsa.090030 (http://www.phytozome.com)
	Cyanophage 9515-10a	ADP00025
	Cyanophage Syn26	ADP00231
	Danio rerio	XP685138
	Daucus carota	AAS15052
	Debaryomyces hansenii CBS767	XP002770286
	Debaryomyces hansenii	CAG89581.2
	Desulfovibrio magneticus RS-1	YP002954416
	Dickeya phage LIMElight	CBW54794
	Dictyostelium discoideum AX4	XP647347
	Drosophila melanogaster	NP608565
	Ectocarpus siliculosus	CBJ30655
	Enterobacteria phage 13a	YP002003938
	Enterobacteria phage 285P	ACV32460
	Enterobacteria phage BA14	YP002003454
	Enterobacteria phage EcoDS1	YP002003741
	Enterobacteria phage Era103	YP001039639
	Enterobacteria phage K1-5	YP654105
	Enterobacteria phage K1E	YP424976
	Enterobacteria phage K1F	YP338094
	Enterobacteria phage K11	P18147
	Enterobacteria phage LKA1	YP001522878
	Enterobacteria phage phiKMV	NP877465
	Enterobacteria phage SP6	NP_853568
	Enterobacteria phage T3	NP523301
	Enterobacteria phage T7	AAA32569

Abb	Taxon/Name/Homolog	Annotation
	Flammulina velutipes	BAB13498
	Fusarium proliferatum	YP001718359
	Gelasinospora sp. G114	O03685
	Glomerella graminicola M1.001	EFQ32004
	Harpegnathos saltator	EFN89497
	Hartmannella vermiformis	ABD62815
	Homo sapiens	AAB58255
HvRpoTm	Hordeum vulgare RpoTm	AJ586899
HvRpoTp	Hordeum vulgare RpoTp	AJ507396
	Klebsiella phage K11	YP002003793
	Klebsiella phage KP32	YP003347522
	Klebsiella phage KP34	YP003347629
	Kluyvera phage Kvp1	YP002308386
	Kluyveromyces lactis NRRL Y-1140	XP453294
	Labrenzia alexandrii DFL-11	ZP05116308
	Laccaria bicolor S238N-H82	XP001880041
	Lachancea thermotolerans	XP002555841
	Leishmania braziliensis MHOM/BR/75/M2904	XP001565382
	Leishmania major strain Friedlin	XP001683631
	Leptosphaeria maculans	CBX89951
	Lodderomyces elongisporus NRRL YB-4239	EDK47326.1
	Lodderomyces elongisporus NRRL YB-4239	XP001526664
	Lolium perenne	CAQ52964
	Magnaporthe oryzae 70-15	XP001402665
	Malassezia globosa CBS 7966	XP001729069
	Meyerozyma guilliermondii	XP001482778
	Meyerozyma guilliermondii ATCC 6260	EDK36161.2
MgRpoTm	Mimulus guttatus RpoTm	mgv1a000798m (http://www.phytozome.com)
MgRpoTmp	Mimulus guttatus RpoTmp	mgv1a000681m (http://www.phytozome.com)
MgRpoTp	Mimulus guttatus RpoTp	mgv1a000925m (http://www.phytozome.com)
MipuRpoT	Micromonas pusilla RpoT	EEH56417.1
	Moniliophthora perniciosa	YP025889
	Monodelphis domestica	XP001375862
	Monosiga brevicollis MX1	XP001746705
	Morganella phage MmP1	YP003324580
MspecRpoT	Micromonas spec RpoT	XP_002503703.1
	Mus musculus	NP766139
	Naegleria gruberi strain NEG-M	XP002670922
NaRpoTm1	Nuphar advena RpoTm1	FN811768
NaRpoTm2	Nuphar advena RpoTm2	FN820498
NaRpoTp	Nuphar advena RpoTp	FN811769
	Nectria haematococca mpVI 77-13-4	XP003048669
	Neosartorya fischeri NRRL 181	XP001261023
	Neurospora crassa	AAA33587
	Neurospora crassa	EAA33334
	Neurospora intermedia	NP049543
NsRpoTm	Nicotiana sylvestris RpoTm	AJ416568
NsRpoTmp	Nicotiana sylvestris RpoTmp	AJ302019
NsRpoTp	Nicotiana sylvestris RpoTp	AJ302020
OlRpoT	Ostreococcus lucimarinus RpoT	ABO98141
OsRpoTm	Oryza sativum RpoTm	AB096014
OsRpoTp	Oryza sativum RpoTp	AB096015
OtRpoT	Ostreococcus tauri RpoT	CAL55557.1
	Paracentrotus lividus	ABM54185
	Paracoccidioides brasiliensis Pb03	EEH18361
	Paramecium tetraurelia strain d4-2	XP001435950
	Pediculus humanus corporis	XP002425302
	Penicillium chrysogenum Wisconsin 54-1255	XP002567345
	Penicillium marneffei ATCC 18224	XP002145390
	Perkinsus marinus ATCC 50983	XP002785982
	Phaeodactylum tricornutum CCAP 1055/1	XP002184121
	Phaeosphaeria nodorum SN15	XP001804112
	Physarum polycephalum	ABB71761
	Phytophthora infestans T30-4	XP002902111
	Pichia kluyveri	CAA72339
	Plasmodium berghei str. ANKA	XP676913
	Pleurotus ostreatus	ABW90146
	Pleurotus ostreatus	YP001504342
	Podospora anserina	S26945
	Polysphondylium pallidum	ABD62816
PotRpoTm1	Populus trichocarpa RpoTm1	EEE88785
PotRpoTm2	Populus trichocarpa RpoTm2	EEF02167
PotRpoTmp	Populus trichocarpa RpoTmp	EEF10826
PotRpoTp1	Populus trichocarpa RpoTp1	EEF03684
PotRpoTp2	Populus trichocarpa RpoTp2	EEE91937
PpRpoTm	Physcomitrella patens RpoTm	in this work
PpRpoTmp1	Physcomitrella patens RpoTmp1	CAC95163
PpRpoTmp2	Physcomitrella patens RpoTmp2	CAC95164
	Prochlorococcus phage P-SSP7	YP214191

Abb	Taxon/Name/Homolog	Annotation
	Pseudomonas phage gh-1	NP813747
	Pseudomonas phage LKD16	YP001522818
	Pseudomonas phage LUZ19	YP001671971
	Pseudomonas phage phi-2	YP003345489
	Pseudomonas phage phikF77	YP002727849
	Pseudomonas phage PT2	YP002117810
	Pseudomonas putida KT2440	NP744415
PtRpoT1	Pinus taeda RpoT1	*unpubl.Beicks et al.*
	Pylaiella littoralis	AAC23956
	Pyramidobacter piscolens W5455	ZP06266917
	Pyrenophora tritici-repentis Pt-1C-BFP	XP001939400
	Ralstonia phage RSB1	YP002213715
	Rattus norvegicus	NP001100236
RcRpoTm	Ricinus communis RpoTm	EEF47677
RcRpoTmp	Ricinus communis RpoTmp	EEF28786
RcRpoTp	Ricinus communis RpoTp	EEF33998
	Rhodopseudomonas palustris DX-1	ZP06359293
	Rosculus sp. ATCC 50888	ABD62817
	Saccharomyces cerevisiae S288c	NP116617
	Salmonella phage phiSG-JL2	YP001949750
	Salmonella phage Vi06	CBV65202
SbRpoT1	Sorghum bicolor RpoT1	XM_002460990
SbRpoTp	Sorghum bicolor RpoTp	XM_002437329
	Scheffersomyces stipitis CBS 6054	ABN67255.2
	Scheffersomyces stipitis CBS 6054	XP001385433
	Schistosoma mansoni	XP002576275
	Schizophyllum commune H4-8	XP003028026
	Schizosaccharomyces pombe 972h-	NP594459
SmRpoTm	Selaginella moellendorfii RpoTm	CAP70041
	Sordaria macrospora	CBI53997
SoRpoTmp	Spinacia oleracea RpoTmp	Y18852
SoRpoTp	Spinacia oleracea RpoTp	Y18853
	Stygamoeba sp. ruffled	ABD62820
	Synechococcus phage P60	NP570316
	Synechococcus phage Syn5	YP001285424
	Synechococcus sp. BL107	ZP01469530
	Taeniopygia guttata	XP002193465
	Talaromyces stipitatus ATCC 10500	XP002487156
TaRpoTm	Triticum avestivum RpoTm	AAF32492
TaRpoTp	Triticum avestivum RpoTp	AAB01085
	Tetrahymena thermophila	XP001013489
	Tetraodon nigroviridis	CAG01374
	Thalassiosira pseudonana CCMP1335	XP002287475
	Theileria annulata strain Ankara	XP953797
	Tigriopus californicus	ABC17826
	Toxoplasma gondii ME49	XP002367014
	Tribolium castaneum	XP967600
	Trypanosoma brucei	AF3035821
	Tuber melanosporum Mel28	XP002841841
	Uncinocarpus reesii 1704	XP002540835
	uncultured phage MedDCM-OCT-S05-C243	ADD95360
	uncultured phage MedDCM-OCT-S08-C41	ADD95501
	uncultured phage MedDCM-OCT-S08-C159	ADD94542
	Ustilago maydis 521	XP758640
	Verticillium albo-atrum VaMs.102	XP003009619
	Vibrio cholerae INDRE 91/1	ZP06028902
	Vibrio parahaemolyticus AN-5034	ZP05891124
	Vibrio phage N4	YP003347903
	Vibrio phage VP93	YP002875649
	Vibriophage VP4	YP249577
	Vitis vinifera	YP002608352
VvRpoTm	Vitis vinifera RpoTm	AM483136
VvRpoTmp	Vitis vinifera RpoTmp	AM488491
VvRpoTp	Vitis vinifera RpoTp	AM453066
	Xanthomonas phage OP1	YP453591
	Xanthomonas phage phiL7	YP002922649
	Xanthomonas phage Xop411	YP001285702
	Xanthomonas phage Xp10	NP858979
	Xenopus laevis	NP001080987
	Yarrowia lipolytica CLIB122	XP499998
	Yersinia pestis phage phiA1122	NP848264
	Yersinia phage Berlin	YP918986
	Yersinia phage phiYeO3-12	NP052071
	Yersinia phage Yepe2	YP002003315
Zea mays	Zea mays	S22768
ZmRpoTm	Zea mays RpoTm	AAD22977
ZmRpoTp	Zea mays RpoTp	AAD22976

6.2 Publications

Richter, U., Kühn, K., Okada, S., Brennicke, A., Weihe, A. and Börner, T. (2010) A mitochondrial rRNA dimethyladenosine methyltransferase in Arabidopsis. *Plant J*, 61, 558-569.

Yin, C., **Richter, U.**, Börner, T. and Weihe, A. (2010) Evolution of plant phage-type RNA polymerases: the genome of the basal angiosperm Nuphar advena encodes two mitochondrial and one plastid phage-type RNA polymerases. *BMC Evol Biol*, 10, 379.

Kühn, K., **Richter, U.**, Meyer, E.H., Delannoy, E., de Longevialle, A.F., O'Toole, N., Börner, T., Millar, A.H., Small, I.D. and Whelan, J. (2009) Phage-type RNA polymerase RPOTmp performs gene-specific transcription in mitochondria of Arabidopsis thaliana. *Plant Cell*, 21, 2762-2779.

Yin, C., **Richter, U.**, Börner, T. and Weihe, A. (2009) Evolution of phage-type RNA polymerases in higher plants: characterization of the single phage-type RNA polymerase gene from Selaginella moellendorffii. *J Mol Evol*, 68, 528-538.

Richter, U., Kiessling, J., Hedtke, B., Decker, E., Reski, R., Börner, T. and Weihe, A. (2002) Two RpoT genes of Physcomitrella patens encode phage-type RNA polymerases with dual targeting to mitochondria and plastids. *Gene*, 290, 95-105.

6.3 Eidesstattliche Erklärung

Hiermit versichere ich, die vorliegende Dissertation eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet zu haben. Die dem Verfahren zugrunde liegende Promotionsordnung ist mir bekannt. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

6.4 Danksagung

Für die Möglichkeit an diesem aussergewöhnlich packenden Thema zu arbeiten und für die wissenschaftliche Betreuung und Kommunikation möchte ich mich bei meinem akademischen Lehrer Herrn Professor Thomas Börner recht herzlich bedanken. Seine grosszügige Unterstützung und wertvollen Anregungen haben nicht nur wesentlich zum Entstehen dieser Arbeit beigetragen, sondern ermöglichten mir auch am wissenschaftlichen Austausch mit Partnerarbeitsgruppen an der Cornell University (USA) und der University Western Australia teilzunehmen. Für die fruchtbare Zusammenarbeit, die anregenden wissenschaftlichen Diskussionen und für ihre Konnivenz bin ich besonders Kristina Kühn dankbar. Meinem Projektleiter Andreas Weihe und besonders Karsten Liere danke ich für vielfältige praktische Ratschläge und den so wichtigen wissenschaftlichen Austausch. Für die phantastische Arbeitsatmosphäre beim täglichen Forschen sei dem gesamten Labor 3 gedankt, danke Steffie, danke Monika, danke Alex, danke Chang und danke Björn. Insbesondere möchte ich mich jedoch beim Guten Geist des Labors – Cornelia Stock - für die nicht enden wollende Unterstützung bedanken. Ein Dankeschön auch an die Arbeitsgruppe von Prof. Reski in Freiburg für die gute Zusammenarbeit. Auch Martin sei besonders erwähnt. Ohne dich ware alles nichts gewesen.

Meinen Eltern bin ich für die allumfassende persönliche Unterstützung, die sie mir zuteilwerden ließen unendlich dankbar. Sie haben ganz wesentlich zum erfolgreichen Verlauf der Arbeit beigetragen. Allen Kollegen und ehemaligen Mitarbeitern der Arbeitsgruppen Genetik und Molekulare Ökologie danke ich für ein herzliches und stimulierendes Arbeitsumfeld.

Der grösste Dank geht allerdings an meine kleine Familie, die oft zu wenig von mir hatte und besonders in den schwierigen Momenten doch so viel Rückhalt für mich bedeutete. Ich liebe Euch und widme meiner Frau und meinen Söhnen Jovan und Paul diese Arbeit.